

AD-780 115

INTERACTIVE SYSTEMS RESEARCH

M. I. Bernstein

System Development Corporation

Prepared for:

Advanced Research Projects Agency

30 April 1974

DISTRIBUTED BY:

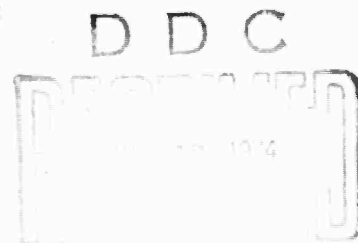
NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

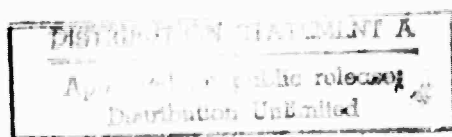
AD780115

**INTERACTIVE SYSTEMS RESEARCH
INTERIM REPORT TO THE DIRECTOR
ADVANCED RESEARCH PROJECTS AGENCY
FOR THE PERIOD
1 OCTOBER 1973 TO 31 MARCH 1974**

30 APRIL 1974



THIS REPORT WAS PRODUCED BY SOC IN PERFORMANCE OF CONTRACT DANC15-73-C-0080
ARPA ORDER NO 2254 PROGRAM CODE NO 3030



Reproduced by
**NATIONAL TECHNICAL
INFORMATION SERVICE**
U S Department of Commerce
Springfield VA 22151

System Development Corporation
2500 Colorado Avenue • Santa Monica, California 90406

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and in-coming annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) System Development Corporation 2500 Colorado Avenue Santa Monica, California 90406		2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	
		2b. GROUP	
3. REPORT TITLE Interactive Systems Research: Interim Report to the Director, Advanced Research Projects Agency, for the Period 1 October 1973 to 31 March 1974			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical			
5. AUTHOR(S) (First name, middle initial, last name) M. I. Bernstein			
6. REPORT DATE 30 April 1974		7a. TOTAL NO. OF PAGES 11, 40 45	7b. NO. OF REFS
8a. CONTRACT OR GRANT NO. DAHC15-73-C-0080		8b. ORIGINATOR'S REPORT NUMBER(S) TM-5243/001/00	
b. PROJECT NO. 3D30			
c.		8d. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d.			
10. DISTRIBUTION STATEMENT Approved for public release; distribution unlimited			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY ARPA/Information Processing Techniques	

13. ABSTRACT

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

**INTERACTIVE SYSTEMS RESEARCH:
INTERIM REPORT TO THE DIRECTOR,
ADVANCED RESEARCH PROJECTS AGENCY,
FOR THE PERIOD
1 OCTOBER 1973 TO 31 MARCH 1974**

30 APRIL 1974

**M. I. BERNSTEIN
(213) 393-9411, EXT 6117**

THIS REPORT WAS PRODUCED BY SDC IN PERFORMANCE OF CONTRACT OANC15-73-C-0080.
ARPA ORDER NO. 2254 PROGRAM CODE NO. 3030

THE VIEWS AND CONCLUSIONS CONTAINED HEREIN ARE THOSE OF THE AUTHOR AND SHOULD
NOT BE INTERPRETED AS NECESSARILY REPRESENTING THE OFFICIAL POLICIES EITHER
EXPRESSED OR IMPLIED OF THE ADVANCED RESEARCH PROJECTS AGENCY OR THE U.S.
GOVERNMENT



System Development Corporation
2500 Colorado Avenue • Santa Monica, California 90406

TABLE OF CONTENTS

	<u>Page</u>
1. INTRODUCTION AND SUMMARY	1
1.1 Speech Understanding Research	1
1.2 Lexical Data Archive	2
1.3 Common Information Structures	2
2. SPEECH UNDERSTANDING RESEARCH	3
2.1 Introduction	3
2.2 Progress and Present Status	3
2.2.1 System Overview	4
2.2.2 The Discourse-Level Controller	6
2.2.3 The Parser	7
2.2.4 The Lexical Matching Procedure	7
2.2.5 The Acoustic-Phonetic Processor	9
2.2.6 System Testing	12
2.2.7 Related Research	14
2.3 Plans	14
2.4 Staff	17
2.5 Documentation and Publications	17
3. LEXICAL DATA ARCHIVE	18
3.1 Introduction	18
3.2 Progress and Present Status	18
3.2.1 Words Analyzed and Sources	20
3.2.2 Sample Semantic Analyses	21
3.2.3 Explanatory Notes on Descriptive Constants	24
3.2.4 Conceptual Analyses of the Notions 'Cause', 'Intentional Action', and 'Responsible'	26
3.2.5 Other Activities	32
3.3 Plans	33
3.4 Staff	33
3.5 Documentation and Publications	33

TABLE OF CONTENTS (Cont'd)

	<u>Page</u>
4. COMMON INFORMATION STRUCTURES	34
4.1 Introduction	34
4.2 Progress and Present Status	36
4.2.1 Common Data Description Language (CDDL) Specification .	37
4.2.2 Types of Data Conversion Functions	38
4.3 Plans	39
4.4 Staff	40

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2-1 Overview of VDMS	5
2-2 The Parser	8
2-3 Lexical Matching Procedure	10
2-4 A-Matrix Processing	13
4-1 Data Flow in the Data Conversion Process	35

1. INTRODUCTION AND SUMMARY

This document is an interim report to the Advanced Research Projects Agency (ARPA) for the period 1 October 1973 through 31 March 1974 on System Development Corporation's (SDC's) research and development program in Interactive Systems Research (ISR). As of the beginning of this period, the program includes three projects: (1) Speech Understanding Research, (2) Lexical Data Archive, and (3) Common Information Structures. The primary emphasis is on Speech Understanding Research and related problems, though the general intent of the overall program is to develop basic technology for improved man-machine interactive systems with application to a variety of anticipated military needs.

The Speech Understanding Research project contains many developments that will be material in enhancing and improving interactive systems by making them more capable and productive when used by the casual user. Of particular importance is the continuing effort to permit such a user to easily and effectively communicate in language forms that are natural to him. In support of this and other work, a project to create a central archive of lexical information for all contractors working in Speech Understanding or other language-based systems has been started.

The world of information processing has been a continuously evolving one since its creation. Of late, the fact that we are becoming ever more dependent on data bases has become obvious (sometimes painfully so). The necessity for the user of an information system to be able to move his data base is as important as a smooth, well engineered interface. Toward this end, the work on minimizing the effort and cost of moving a data base from one system to another is continuing.

The following paragraphs summarize the project activities for the past half year.

1.1 SPEECH UNDERSTANDING RESEARCH

The Speech Understanding Research (SUR) project continues to progress toward its goal of creating a demonstrable prototype Voice-controlled Data Management System (VDMS) using free-form spoken English as input. The two parallel development efforts of last year in acoustic-phonetic processing and linguistic processing were merged into a single system that was demonstrated to the ARPA SUR Group Review Team in December 1973. Work has continued on all components of the system, with the objective of both increasing the vocabulary and loosening the grammar and syntax of the language accepted by the system. The improved version of VDMS should be demonstrable in the fall of 1974.

1.2 LEXICAL DATA ARCHIVE

The Lexical Data Archive project began this past October with the objective of providing a centrally collected body of lexical information on the union of lexicons used by the ARPA SUR Contractors in their collective and individual endeavors. Toward this end, the Semantically Oriented Lexical Archive (SOLAR) has been designed, and much of the relevant data has been either located or acquired. Several SOLAR files already exist, and others are being created. Information from these existing files is available to its using community manually at present. Several examples of files contained in SOLAR are presented. Within the next six months, more extensive information will be available, most of it automatically.

1.3 COMMON INFORMATION STRUCTURES

Work is continuing on creating the methodology and necessary components of a system for moving data bases from one system to another at minimal effort and cost. Current conventional methods have been studied, and a unique approach has been devised that makes maximal use of the Data Management System tools by which data bases are created, updated, and queried. Three descriptive languages needed to describe the transfer process have been defined. They are a Common Data Description Language (CDDL), a Common Data Translation Language (CDTL), and a Common Data Format Language (CDFL). A first version of the CDDL is presented. Work will continue on the creation of the other two languages, and, eventually, the method will be implemented and tested.

2. SPEECH UNDERSTANDING RESEARCH

2.1 INTRODUCTION

The continuing long-term goal of the SDC Speech Understanding Research (SUR) project is to develop and implement a data management system that is controlled and operated by its users through free-form spoken English. The basic approach taken to achieve this goal is distinguished by a modular system architecture that embodies phonological and linguistic processes and an acoustic-phonetic processor. The system architecture enables a complete assembly of multidirectional parsing processes to operate in parallel on the same or different segments of an input utterance. Two major advantages are obtained from this: (1) the system may start working on the least ambiguous portions of the input, and (2) predictions need not be limited to near neighbors of a recognized input segment but may be applied to any portion of the entire utterance.

The acoustic-phonetic processor contains the processes that extract acoustic information from the speech signal and make acoustic-phonetic labeling decisions. The processor we are developing reflects the fact that the speech signal is never wholly unambiguous; any attempt to precisely label phones and their boundaries must recognize and allow for this ambiguity in mapping the extremely large number of speech sounds into the relatively small set of acoustic-phonetic transcription symbols. Accordingly, in this processor, each acoustic-phonetic segment is multiply labeled, and each label is assigned a score. Scores are based on a measure function that is, in turn, based on feature parameters previously developed for each speaker.

As a first step toward the long-term goal, we decided to construct and refine a limited voice-controlled data management system (VDMS) that could accept continuous speech and be demonstrably usable by at least two speakers. Within this system, limitations were planned with respect to both the vocabulary and the syntax of the English subset permitted.

2.2 PROGRESS AND PRESENT STATUS

At the beginning of the present contract year, two separate versions of VDMS had been constructed and tested: Version A, which operated on the SUR laboratory Raytheon 704 mini-computer in conjunction with an IBM 370/145, and which incorporated the modular system architecture, and Version B, which operated entirely on the Raytheon 704, and which embodied the multiple-labeling philosophy of the acoustic-phonetic processor described above. Both versions allowed the user to access a data base of information about the length, beam, draft, armament, and other characteristics of submarines in the naval fleets of the United States, Soviet Union, and United Kingdom. The total vocabulary of each system was approximately 150 words. The query language used to access this information could be described by about 35 syntax

equations for Version A and about 20 syntax equations for Version B; the difference is accounted for by the fact that Version A contained report generation capabilities that Version B did not. Typical queries that could be accommodated by either system are:

"Total quantity where type equals nuclear and country equals USA."

"Print type where missiles greater than seven."

Both of these initial versions of VDMS had been tested with a large number of utterances and had achieved reasonably good results. The first major task undertaken during this contract year was the construction of a single full-scale version of VDMS that combines the best elements of the two versions. This new version of VDMS (Ritea, [1]) was completed and successfully demonstrated in late 1973. Its major characteristics are described in this section.

2.2.1 System Overview

The overall configuration of VDMS is characterized by three major processing modules:

- 1) The linguistic processor, which contains the parser and a discourse-level controller;
- 2) The acoustic-phonetic processor, whose results are contained in an array of data called the A-matrix;
- 3) The lexical matching procedure, which performs matches of predicted words at the syllable level, using various applications of phonological rules to assist in its matchings.

The pattern of communication among these modules is illustrated in Figure 2-1. The speech from the user is input to the acoustic-phonetic processor, which forms an array of acoustic-phonetic data for use by the parser. At the beginning of the processing of an utterance, the discourse-level controller provides a variety of predictions and restrictions on what is allowable or expected in this utterance. The predicted words are transmitted to the lexical matching procedure, which looks for the words in the acoustic-phonetic data. The parser and lexical matching procedure then pass predictions and verifications back and forth to one another in an effort to understand the utterance. Once it is understood, the utterance is passed to the data management system, which forms an appropriate response. The response is then passed to the discourse-level controller and to the user. The discourse-level controller is thus updated to aid in future predictions.

The logical flow of control and data between all of the modules is specified by a language unique to VDMS, called the Control Structure Language (CSL). Using CSL for program control, new modules may be implemented, and data paths among

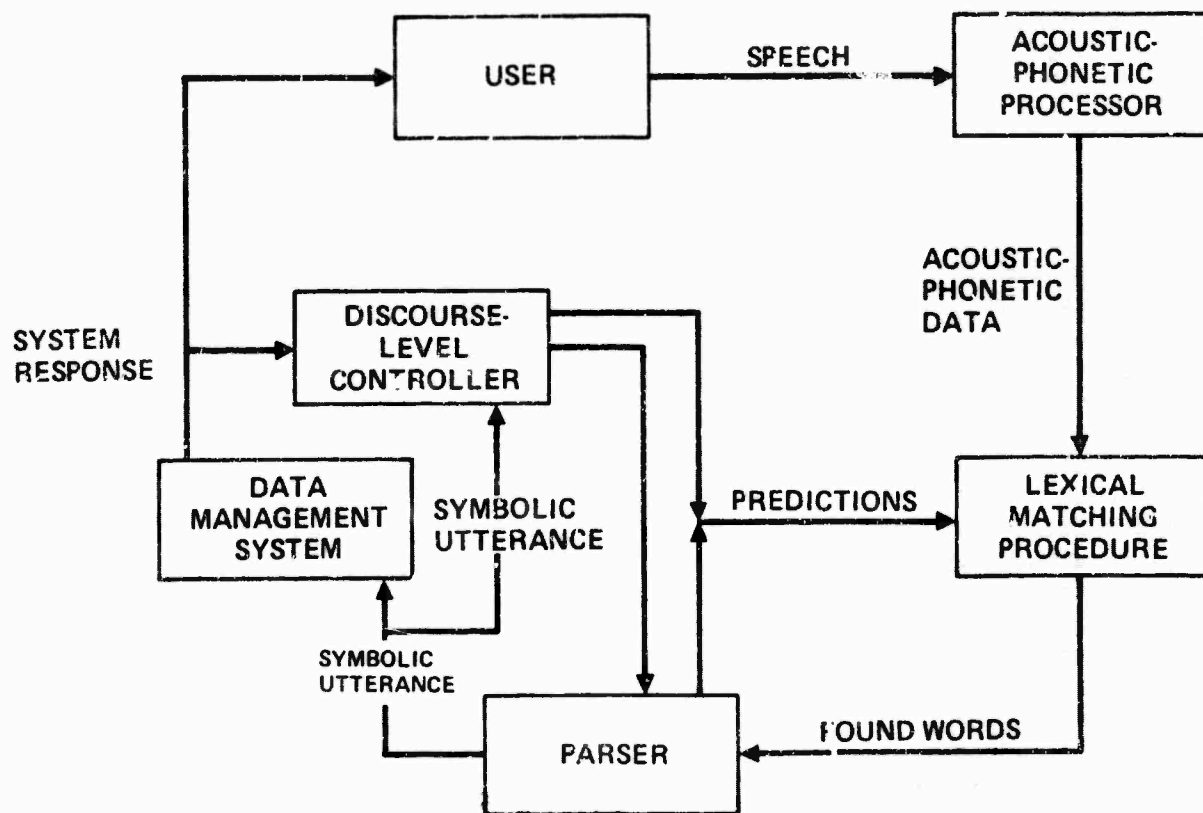


Figure 2-1. Overview of VDMS

modules in VDMS may be modified without major reprogramming of the system. In addition, CSL has the following features:

- 1) It allows for logical parallel execution of modules.
- 2) It provides for the running of modules on remote computers.
- 3) Using a trace and debugging provision, breakpoints can be inserted for monitoring the flow of data through the system.
- 4) Changes in the order of execution of the various modules may be specified.
- 5) Data dependencies among modules may be controlled.

A more detailed discussion of CSL is given by Barnett [2].

2.2.2 The Discourse-Level Controller

The discourse-level controller comprises two modules: the user model and the thematic memory. The user model determines what query "state" the user is in and predicts the kinds of grammar that may appear in his next interaction with the system. Some sample "states" are system login, interactive query mode, report generation mode, and user aids. If the user is in interactive query mode, the user model will predict syntax equations for the next interaction, such as those for Print, Repeat, Count, Subset, or Total, each of which is the first word of an interactive query statement. The words Explain and Describe are the first words of typical user-aid commands. Each prediction carries with it a confidence level such that the higher the confidence level, the more liberal the system will be in overlooking errors in recognition.

The thematic memory is concerned with particular content words that might occur in the next utterance; it is not concerned with syntactic terminals such as the digits or the word "Print." Several pieces of information are kept about each word as it is used: e.g., how long (how many utterances ago) it has been since the word was used and how likely it is that the word will re-occur, depending on how it was used originally. For example, if the user said "Show category," the assumption is that the next command will probably involve something about the categories of submarines in the data base.

In addition to looking for content words in user commands, the thematic memory also keeps a record of any non-numeric symbolic responses from the data management system. These responses are also used to predict words that are highly likely to occur in the next utterance.

Throughout a dialogue, the various content words as predicted by the thematic memory are "aged" from utterance to utterance, and their likelihood of being used is diminished if they have not been reused. If the "Confidence of Prediction" drops below a threshold, then the word is removed from the thematic memory and dropped from consideration until it is used again. Also, if duplicate entries occur within an utterance, then the age and original merit are modified to take care of this effect.

2.2.3 The Parser

The basic linguistic unit used for the parsing strategy in VDMS is the phrase, which consists of one or more vocabulary words (up to the complete utterance) linked together in a syntactically and semantically correct order. Some examples of phrases are "country and category" and "quantity equals five." The parser attempts to predict phrases using the user model, thematic-patterning, and grammatical and semantic constraints information provided by the discourse-level controller. Predicted phrases are matched against the acoustic-phonetic data for acceptance or rejection. Accepted phrases are then concatenated to form a larger phrase, which is then analyzed to see whether it is a complete utterance.

The parser consists of four major modules:

- 1) The classifier
- 2) The bottom driver
- 3) The top driver
- 4) The side driver

The classifier's task is to assign a syntactic category to each word accepted by the lexical matching procedure. Some typical syntactic categories and examples are:

Item name ("country")
Item value ("US?")
Syntactic terminal ("print").

A syntactic category, as generated by the classifier, is used by the other modules of the parser to generate predictions about allowable syntax in other parts of the utterance. The bottom driver is a typical "bottom-up" module, which takes found phrases and determines how they may be used in completing the parsing of a complete utterance. The top driver takes predicted phrases and from them derives either a syntactic terminal or a shorter phrase to be looked for next. The syntactic terminals are sent to the lexical matching procedure, which then attempts to match each one against the acoustic-phonetic data. The side driver takes completed or partially completed phrases from the bottom driver. If a phrase is incomplete, the side driver determines which part to look for next, and it will ask the top driver to locate the missing part. On the other hand, if the phrase has been completed, the side driver analyzes it to see if it is a legal complete utterance. If it is, the side driver terminates the parsing activities of all modules and transmits the symbolic form of the hypothesized utterance to the data management system and the discourse-level controller. Other completed phrases (which do not cover the entire utterance) are used to "bottom-drive" the entire system or "bottom-drive" that partial parse up one level to create larger, more complete phrases. The flow of processing within the parser is shown in Figure 2-1.

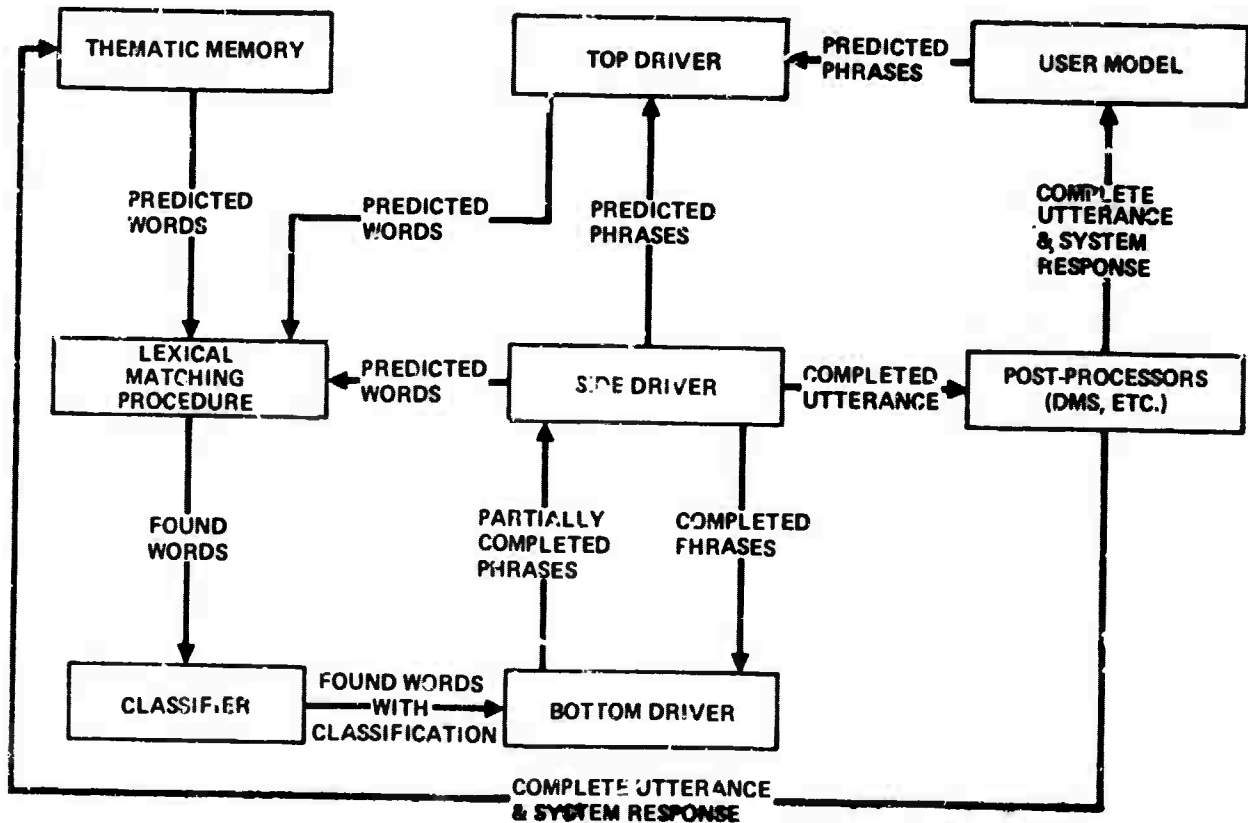


Figure 2-2. The Parser

2.2.4 The Lexical Matching Procedure

The lexical matching procedure verifies or rejects a predicted word through pattern-matching against the available acoustic-phonetic data. A detailed description of the procedure is given by Weeks [3].

The syllable is the unit that is used in the lexical matching process. The linguistic issues concerning its existence or form have been sidestepped by giving it an algorithmic definition. All words have syllable divisions marked in the lexicon, and some of the phonemic rules are written in terms of these boundaries. Within a syllable, most of the co-articulation is internal; effects over boundaries are handled separately. Because a good percentage of phonetic dependencies occur within these units, rules can be conveniently applied. Under this approach, a separate set of rules must be set up for dealing with interactions over boundaries. Word boundaries can then be considered as special cases of syllable boundaries.

Figure 2-3 is a block diagram of the lexical matching procedure. When a word is predicted in orthographic form, its phonemic representation is extracted from the lexicon. A set of phonemic rules is applied to the phonemic representation to obtain a set of lexical variants. These variants arise by phonemic replacements as dictated by the rules. The set of lexical variants is then sent to the main matching procedure, where each is matched one by one against the acoustic-phonetic data on a syllable-by-syllable basis. The boundary analysis is done in conjunction with the syllable matching and attempts to compensate for articulation across syllable and word boundaries. The resulting scores for each lexical variant are then sent to the main matching procedure, which decides which possibility gives the best over-all score. The structure of the phonemic rules pass is described by Barnett [4].

2.2.5 The Acoustic-Phonetic Processor

Speech is input interactively in a very quiet sound booth (signal-to-noise ratio better than 50 dB) using a Sony ECM-377 condenser microphone, which has an essentially flat frequency response to beyond 10,000 Hz. Low-level preemphasis is employed, shaping the frequency response with a zero at 300 Hz and a pole at 3,000 Hz. This speech signal is bandlimited to 9,000 Hz and digitized at 20,000 samples per second using a 12-bit analog-to-digital converter. The input speech is saved directly on digital media with no intervening analog recording steps.

The digitized speech is then passed through a digital-to-analog converter, and the resulting analog waveform is passed through three hardware filters having nominal bandpasses of 150 to 900 Hz, 900 to 2,200 Hz, and 2,200 to 5,000 Hz, respectively. Over 10-msec. intervals, two parameters are extracted from each of the three filter outputs: the maximum peak-to-peak amplitude and a count of zero crossings. The resulting six parameters (two from each of the

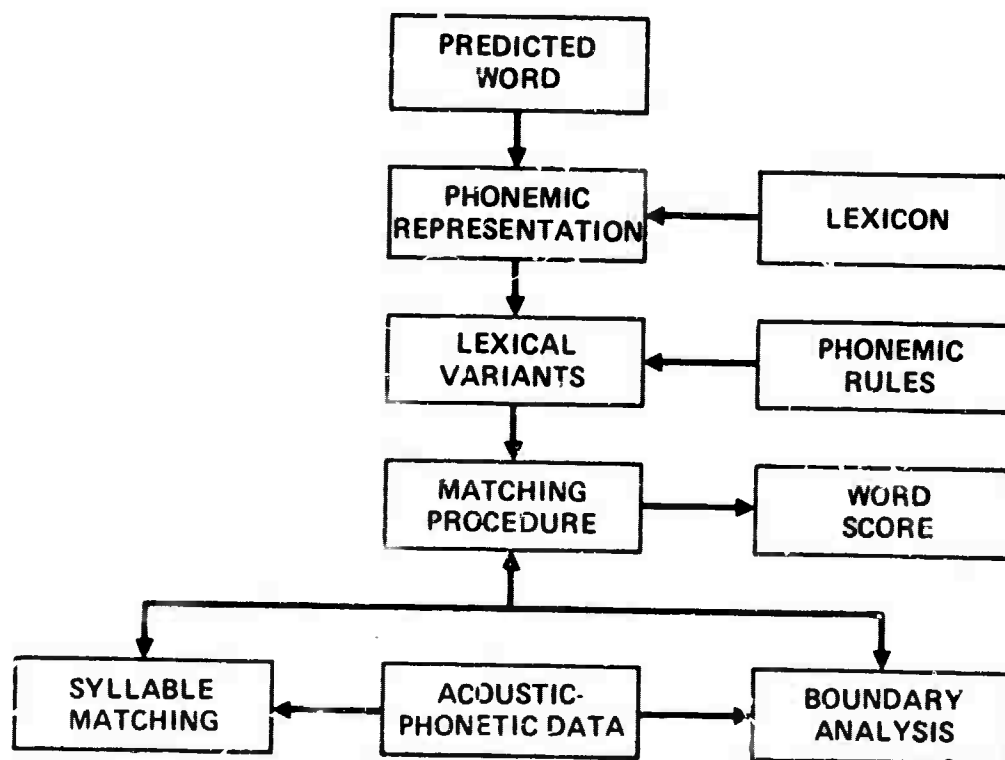


Figure 2-3. Lexical Matching Procedure

three filtered signals) are used to assign a rough acoustic label to each 10-msec. segment. Five labels are currently used: VW (vowel-like), SS (strong frication), SI (silence), UV (low-amplitude voiced or unvoiced), VC (all other--usually weak voicing). The next step in the processing is to refine these rough labels, i.e., impose on each 10-msec. segment a more accurate label than the five above.

Within the classes VW and VC, more specific labels are assigned using a vowel-recognition strategy based on speaker-dependent vowel formant information. This information is compared with the formants (obtained with the use of a Linear Predictive Coefficient (LPC) spectrum) at selected instants in time for each vowel to be identified. A modified Euclidean distance function is used to compute the relative distances between the candidate formant values and the pre-stored speaker-dependent vowel formant values. The closest three vowels are selected, and associated scores are assigned to these choices based on the values of the distance function.

Fricatives and plosives are characteristically found within sequences of segments labeled SS, VC, or UV. For these areas, a technique called the Low-Coefficient LPC (LCLPC), described by Molho [5], has been shown to provide meaningful spectra that correspond well with both acoustic-phonetic theory and with the experimental results of others. Time resolution is sufficiently narrow to allow independent spectral analysis of the release, frication, and aspiration portions of an unvoiced plosive or to demonstrate spectral change within a consonant cluster, so that clusters such as /ks/ and /ts/ may often be distinguished. For analysis of unvoiced speech, the LCLPC uses the auto-correlation method with eight coefficients and a 6-msec. Hamming window. Analysis of spectra obtained in this way allows the following five classes to be distinguished:

- 1) labial or dental (LD)
- 2) alveolar (AL)
- 3) alveopalatal (AP)
- 4) palatal or velar (PV)
- 5) voiced or low energy (VS)

These classes correspond roughly to the spectral characteristics of unvoiced fricatives and plosives. Moreover, there is a correspondence between these classes and the articulatory positions of unvoiced fricatives and plosives. The classes LD, AL, AP, and PV ideally contain the following phonemes:

LD: /p/, /f/, /θ/
AL: /t/, /s/
AP: /ʃ/
PV: /k/

Experimentation has also confirmed that the glide /w/ and the liquid /l/ characteristically occur within the VW or VC classes. The present approach to recognizing these phonemes is to augment a speaker's vowel formant table with the formant frequency values for /w/ and /l/. These formant values have consistently been easily distinguishable from the formants of the vowels and have enabled the system to accurately isolate and recognize /w/ and /l/. The glide /y/ and the liquid /r/ are handled indirectly, again with the use of the speaker-dependent vowel formant table: if a 10-msec. segment has been labeled /i/, it is assumed that the segment could be a /y/ with equal probability, and both labels are then assigned to the segment with the same score. If a segment has been labeled /ɜ/ (again with the aid of the vowel table), the label /r/ is assigned to the same segment with an equal score.

Although the system is not yet able to distinguish the various elements within the class of nasals, viz., /m/, /n/, /ɲ/, /m/, /n/, a single class name (NA) is used and has proved quite reliable. A segment is labeled NA based upon some simple tests involving the amplitudes and bandwidths of F_1 , F_2 , and F_3 . All of the aforementioned segment labeling procedures are used to construct an array of acoustic-phonetic data called the A-matrix. The construction of the A-matrix is shown in Figure 2-4. Each row of the A-matrix corresponds to a 10-msec. segment of speech and contains a rough segment label (VW, SS, SI, VC, or UV), one or more refined segment labels and associated scores based on the above procedures, formant frequency values, and estimates of fundamental frequency, RMS energy, and other acoustic-phonetic parameters used in the assignment of the phoneme and phoneme-class labels.

2.2.6 System Testing

In the present configuration, speech is digitally recorded and saved on disk, as described above, using a Raytheon 704 computer. The same computer then creates an A-matrix from the digitized waveform. The A-matrix is then sent (via a direct hardware link) to an IBM 370/145 computer, which then performs all subsequent linguistic processing of the utterance and returns a response to the user. Since the system is dependent upon thematic patterning to aid in the understanding of an utterance, it is necessary for the user to interact with VDMS using goal-directed dialogues. For purposes of testing, ten dialogues were created with an average of ten utterances per dialogue. Each of two male speakers (for whom vowel formant tables had been previously constructed) recited the sets of dialogues. In this initial test of VDMS, an average of 52% of the utterances were correctly understood. Analysis of these preliminary results has shown that this figure can be increased by correcting program errors and implementing some modifications to the phonological processes and lexical matching procedure.

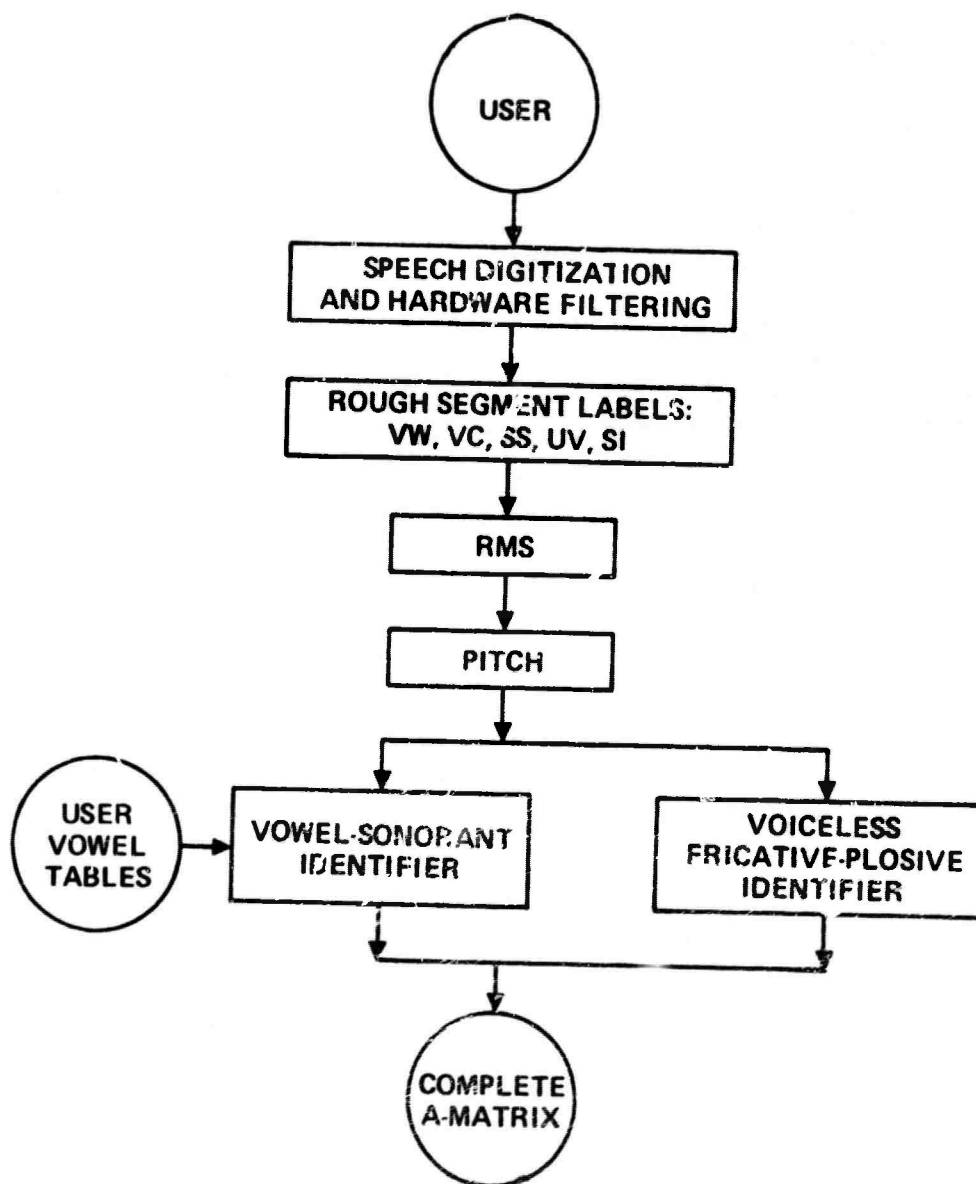


Figure 2-4. A-Matrix Processing

2.2.7 Related Research

An experiment was designed to compare the formant 1 (F1) and formant 2 (F2) frequency movements of vowels next to /r/ with the same vowels before other consonants [6]. Lehisto's data¹ (obtained from spectrograms) on the vowel allophones associated with /r/ were used for comparison purposes. The data for this experiment were based on formant trajectories computed by LPC techniques on the Raytheon 704 computer.

The results of this experiment confirmed Lehisto's work, which indicates that there is a change in some vowels in a retroflexed environment. The change in vowels after /r/ is minimal except for /i/, but the change in vowels before /r/ is considerable.

This was a preliminary experiment in which the number of subjects and samples was small. However, the results can be used to develop a retroflexed vowel space on the basis of a non-retroflexed vowel space and to compare the identification of vowels using this new F1-F2 space with the identification of vowels using the non-retroflexed F1-F2 space.

An algorithm was designed to automatically distinguish the nasals /n/, /m/, and /ŋ/ from each other [7]. Spectral analysis is performed on the Raytheon computer using an LPC model to locate the formants of these phonemes. By comparing the formant frequencies of unknown nasals to prototype values derived from normalization utterances, the algorithm was able to correctly identify nasals in 72% of the cases tested. Experimentation has indicated that (1) automatic techniques can be employed to distinguish nasals in continuous speech; (2) linear prediction can be used to effectively analyze the spectra of these phonemes; and (3) speaker-dependent tables of prototype nasal formants extend these results to multiple-speaker environments.

2.3 PLANS

The original goal to be reached by the end of the current contract year (as specified in SDC Proposal 73-5674) consisted of enlarging the vocabulary of the query language and loosening the grammar to make the language easier and more natural to use. Specifically, VDMS was to contain a vocabulary of about 500-600 words, and the grammar was to be modified to admit the following capabilities:

- 1) More natural form of expression for integers (example: "thirty-four" instead of "three four").
- 2) Facility for inter-item comparisons (example: "Print category where surface speed greater than submerged speed").
- 3) Use of strings of inequalities (example: "Print type where draft greater than seven and less than nine").

¹I. Lehisto, Acoustical Characteristics of Selected English Consonants, Mouton & Company, The Hague, 1964.

- 4) Simple arithmetic calculations (example: "Print category where surface speed greater than three times submerged speed").
- 5) Interrogative sentences.

However, recent discussions and negotiations (at the request of ARPA) with the Stanford Research Institute (SRI) have yielded a cooperative operating plan in which this original goal has been changed. We believe that our proposed modifications for September 1974 will provide a stronger base from which to proceed to our long range goal, viz., the five-year system as characterized by Newell, et al.¹ The modifications are summarized below and described in SDC's current proposal to ARPA (SDC Proposal 74-5490, 15 April 1974).

Two major systems are planned for the end of the current contract year (September 1974):

- 1) Version A, which will have a vocabulary of about 150 words, an English-like grammar (with the aid of the SRI parser), and all other features similar to the present VDMS.
- 2) Version B, which will be the present VDMS with a vocabulary of about 300 words.

Version A will provide the first step toward a natural English VDMS. The main use of Version B will be to test the robustness of the acoustic-phonetic algorithms on a larger vocabulary.

The following tasks are currently in process to aid in the construction of Version A:

- 1) System Architecture--The design concepts underlying the system developed by SDC and SRI during the first two years of the ARPA program were sufficiently similar that it seemed reasonable to develop a single system that would combine the procedure for acoustic, phonetic, and phonological analysis developed by SDC with the procedures for syntactic, semantic, and pragmatic analysis developed by SRI. To accomplish this goal, we are re-analyzing the design concepts of each and developing a new system design. (Responsibility: SDC, SRI)
- 2) Parsing--The design of the SRI parser is being revised to accommodate the SDC procedures for acoustic-phonetic analysis and lexical mapping. Changes also may result from additions to the grammar. An initial version of the revised parser will be operating in the new system by September. (Responsibility: SRI)

¹ Speech-Understanding Systems: Final Report of a Study Group.
Pittsburgh, Carnegie-Mellon University, 1971.

- 3) Grammar--A major reorganization of the SRI grammar is being made. The rules included in the initial revision will be influenced by dialogues collected during the early protocol experiments (see below). The structure of the grammar is being changed so that additions and modifications can be made more easily. (Responsibility: SRI)
- 4) System Software--Additions to the system software, particularly extensions and modifications of SDC LISP, will be made as required. (Responsibility: SDC)
- 5) System Hardware--Work is proceeding on the ARPANET interface for the SDC computer facility. As the special-purpose signal processing and acoustic analysis computers are acquired, they will be interfaced into the system. (Responsibility: SDC)
- 6) Lexical Matching Procedure--Substantial changes will be made for coordination with the SRI parser and grammar for Version A. (Responsibility: SDC)
- 7) Protocol Experiments--Persons are being located who are thoroughly familiar with information retrieval operations on data of the type contained in the submarine task domain. On the basis of interviews with them, a set of tasks will be identified and used in eliciting task-oriented dialogues. Early results will be used immediately to determine additions to the vocabulary and to the data base. Subsequent analyses will guide revisions to the grammar and the design of an effective discourse model. (Responsibility: SRI)
- 8) Semantics--Major changes in the semantics developed for the original SRI task domain will be made, initially to accommodate the data management task. Subsequently, as the data base is extended in accord with the findings from the protocol experiments, additional modifications will be made. (Responsibility: SRI)

The following tasks are currently in process to aid in the construction of Version B:

- 1) Acoustic-Phonetic Analysis--The algorithms used to build the A-matrix for an utterance are being refined and extended, and new procedures are being added to improve the accuracy of the classifications. (Responsibility: SDC)
- 2) Lexical Matching Procedure--The current lexical matching procedure in VDMS, with minor modifications, will be adequate for testing the system with the extended vocabulary. (Responsibility: SDC)

2.4 STAFF

H. B. Ritea, Project Leader
J. Barnett
W. A. Brackenridge
R. A. Gillmann
I. Kameny
P. Ladefoged (Consultant)
L. Molho
D. Pintar
R. V. Weeks

2.5 DOCUMENTATION AND PUBLICATIONS

- [1] H. B. Ritea, "A Voiced-Controlled Data Management System," Proceedings of the IEEE Symposium on Speech Recognition, Pittsburgh, Pa., April 15-19, 1974.
- [2] J. Barnett, "Module Linkage and Communication in Large Systems," Proceedings of the IEEE Symposium on Speech Recognition, Pittsburgh, Pa., April 15-19, 1974. (Companion Volume of Invited Papers, in press.)
- [3] R. V. Weeks, "Predictive Syllable Mapping in a Continuous Speech Understanding System," Proceedings of the IEEE Symposium on Speech Recognition, Pittsburgh, Pa., April 15-19, 1974.
- [4] J. Barnett, "A Phonological Rule Compiler," Proceedings of the IEEE Symposium on Speech Recognition, Pittsburgh, Pa., April 15-19, 1974.
- [5] L. Molho, "Automatic Recognition of Fricatives and Plosives in Continuous Speech," Proceedings of the IEEE Symposium on Speech Recognition, Pittsburgh, Pa., April 15-19, 1974.
- [6] I. Kameny, "Comparison of the Formant Spaces of Retroflexed and Non-retroflexed Vowels," Proceedings of the IEEE Symposium on Speech Recognition, Pittsburgh, Pa., April 15-19, 1974.
- [7] R. A. Gillmann, "Automatic Recognition of Nasal Phonemes," Proceedings of the IEEE Symposium on Speech Recognition, Pittsburgh, Pa., April 15-19, 1974.

3. LEXICAL DATA ARCHIVE

3.1 INTRODUCTION

The Lexical Data Archive (LDA) project has addressed itself to the task of providing the ARPA SUR project contractors with semantic and syntactic data for the words in their lexicons. Being devoted exclusively to lexical research, LDA can assure a broad range of services for each SUR project without the triplication of effort and resources which would be required if the three SUR projects were to perform the same tasks. LDA will monitor a broad range of lexical data sources, select the data having potential payoff for speech understanding, format that data for archiving purposes, and provide for its dissemination to the appropriate SUR projects.

3.2 PROGRESS AND PRESENT STATUS

Since the initiation of LDA in September 1973, the following achievements have been realized. First, the design of the Semantically-Oriented Lexical Archive (SOLAR) has been completed. This task included deciding what types of lexical data to collect, determining the data collection procedures, writing the specifications of the programs needed to extract data from machine transcripts, and designing the logical structure of the files to be built. In accordance with questionnaire responses from the SUR projects, SOLAR will consist of eight files.

- 1) A word index will allow a user to easily determine the types of data available for a given word.
- 2) A bibliographic reference file can be used in conjunction with other files, to allow abbreviated references and can also be used as a separate resource for guidance into the literature.
- 3) A file of semantic analyses will contain formal treatments of the semantic properties of individual words as found in the literature. Notes will also be provided explaining the descriptive constants employed in these analyses.
- 4) Integrated summaries of analyses given in the philosophical literature for concepts central to the descriptive constants used in the semantic analyses will be constructed.
- 5) Collocational information found in the definitions of Webster's Seventh New Collegiate Dictionary (W7) is to be machine-extracted and made accessible via the words to which it pertains.
- 6) Definitional expansions controlled by the words within a particular SUR lexicon will be constructed.

- 7) Semantic fields will be built for each SUR word by tying to them the words found in certain definitional and synonymitive relationships in W7 and Roget's International Thesaurus (Roget).
- 8) Finally, every context of each SUR word in the W7 definitions and in the Brown Corpus will be entered in a key-word-in-context (KWIC) file.

For a more detailed discussion of the contents of each of these files see Diller and Olney [1].

A second achievement has been the collection of a significant amount of data for the first three files just described. All of the SUR words in use as of about January 1974 have been entered into the word index, together with their W7 parts of speech and an indication of the SOLAR data available for each. About 2000 references to linguistic and philosophical documents have been entered into the bibliographic file. Indexing by author, title, and keyword (among other parameters) is possible for each bibliographic entry. Approximately 70 semantic analyses have been entered into the third file. A list of the words to which these analyses apply is given in Section 3.2.1. Section 3.2.2 presents four semantic analyses of 'give' as they currently appear in the archive. In Section 3.2.3 notes on the descriptive constants used in one of those analyses are given. Without such notes, some of the analyses would be virtually incomprehensible (e.g., the first one, which is taken from Bendix's dissertation). Section 3.2.4 presents three conceptual analyses, two of which are especially relevant to descriptive constants used in Bendix's semantic analysis of 'give', viz., 'cause(s)' and the sequence '(C AN-F)'. Bendix's comments on these constants (see our explanatory notes E-113, E-115, and E-118 in Section 3.2.3) include the following reservations: "Ultimately we would want to define the metalinguistic term 'cause' more precisely. This would be intended as a definition not of causation but of what we claim a speaker is asserting when he uses a form whose meaning we define as containing the component 'cause'." While we certainly grant that the task of defining Bendix's metalinguistic term 'cause' is distinguishable from the task of defining causation, we question whether the former task can be accomplished successfully in total isolation from the latter. Bendix's use of 'cause' as a metalinguistic term in his semantic analysis of 'give' is typical of the recent practice of linguists in that the meaning he defines for that term corresponds closely to a sense of the English word 'cause' (see p. 29 of [1]). Under these circumstances the attempt to make the conceptual content of the metalinguistic term more precise almost inevitably leads into an investigation of the conceptual content of the corresponding word sense, if only for the sake of differentiating them more sharply.

As with a great many of the metalinguistic terms used by linguists for semantic analysis, the conceptual content of the word sense corresponding to Bendix's 'cause' has been investigated extensively by analytic philosophers. It is noteworthy in this connection that Bendix's semantic analysis of 'give' suffers

from a defect that is highlighted by philosophical analyses of 'cause' and the closely related notion 'intentional action'. The essence of Bendix's analysis of 'give' is the claim that 'C gives B to A' is equivalent to 'C causes (A has B)'. However, it is certainly false, if C gives B to A, that any kind of causal connection between something C does and A's subsequently having B will suffice. Clearly what C must do is perform an intentional action which, C anticipates, will result in A's having B (see the integrated summary of conceptual analyses of 'intentional action' (I-103) for details). In other words, Bendix's semantic analysis of 'give' should be augmented by a clause containing a descriptive constant (metalinguistic term) that invokes a notion of intentional action. In addition to making it easier for linguists to recognize such omissions in their semantic descriptions, our integrated summaries of conceptual analyses will give them a head start in working out adequate definitions of the descriptive constants they decide to use. The complexity illustrated by the summary we give for 'intentional action' is typical of the conceptual content of basic, commonplace notions into which many word meanings seem readily decomposable. There is a striking parallel between the ease with which a person can use such notions appropriately in thinking and communicating without being able to specify their conceptual content in detail and the ease with which he can utter grammatically correct sentences without being able to specify their grammatical structure in detail.

3.2.1 Words Analyzed and Sources

Semantic analyses have been entered for the 50 words following. Since some words have more than one analysis, approximately 70 analyses have been completed for these 50 words.

all, and, any, assemble, bear, bring, carry, come, conduct, cross, descend, disperse, drive, drop, eight, emerge, enter, escape, excuse, five, flow, four, get, give, go, head, lead, leak, may, move, navigate, nine, part, penetrate, proceed, quit, sail, scale, scatter, separate, seven, six, slip, some, start, submerge, take, three, turn, two

The analyses are taken from the following sources:

Bendix, Edward (1966). Componential Analysis of General Vocabulary: The Semantic Structure of a Set of Verbs in English, Hindi, and Japanese. Bloomington, Indiana: Indiana University Press.

Dixon, R. M. W. (1973). "The Semantics of Giving," in The Formal Analysis of Natural Languages. Ed. by Gross, Maurice et al. The Hague: Mouton.

Fillmore, Charles (1969). "Verbs of Judging: An Exercise in Semantic Description," in Papers in Linguistics 1.1, pp. 91-117.

Ikegami, Yoshihiko (1969). The Semiological Structure of the English Verbs of Motion.

Blackendoff, Ray (1971). "Modal Structure in Semantic Representation," Linguistic Inquiry 2.4, pp. 479-514.

Katz, Jerrold (1972). Semantic Theory. New York: Harper and Row.

Staal, J. F. (1968). "And," Journal of Linguistics, Vol. 4, pp. 79-81.

It should be emphasized that the sources do not reflect a bias that will be maintained in the project. An exceedingly large number of references relevant to the SUR words have been found, and they will be worked through systematically, beginning with those containing words occurring in most of the SUR lexicons.

3.2.2 Sample Semantic Analyses

Four semantic analyses have been entered into the archive for the word 'give'. These have been taken from three separate sources.

SOLAR Word: GIVE

Domain: AP

SMALLWORDS

VOCAB

Semanal #: S-102

Source: BENDIX, EDWARD [1966]

Accompanying Words: GET, TAKE, LOSE, KEEP, LEND, GET RID OF, BORROW, FIND, BE, HAVE

Author's Sense: NO RESTRICTION OF THE SENSE IS INDICATED

W7 Sense: HOMOGRAPH 1, VT (SENSES 1 - 15)

Prelim. Qual.: BENDIX IS CONCERNED WITH THE PARADIGM "C GIVES A B". HE PROCEEDS ON THE HYPOTHESIS THAT 'GIVE' SIGNIFIES A CHANGE OF STATE INVOLVING 'HAVING' AND THEREFORE INCLUDES THE COMPONENT 'A HAS B'.

Components: 'C CAUSES (B AN-RH A)' = 'C CAUSES A TO HAVE B' = 'C CAUS (A HAS B)'

Entailments: (NOW) A HAS B

Informal Explic.: "...WHEN A SPEAKER ASSERTS 'C GIVES A B', WE CLAIM HE IS ASSERTING THAT THERE IS A CAUSAL CONNECTION FROM SOMETHING 'C' DOES OR IS ('C CAUSES' = '(C AN-P) CAUSES') TO THE ASSERTION 'A HAS B'." [pp. 69-70]

Final Qualif.: "...THERE IS NO COMPONENT [INDICATING] THAT 'C' HAS 'B' BEFORE GIVING IT TO 'A'..." [p. 70] "NOTE ALSO THAT THE DEFINITION IS NOT PHRASED AS 'C CAUSES (A GETS B)'." [p. 70]

SOLAR Comments: THE TERM 'NOW' IS USED BY BENDIX TO INDICATE THE CHANGE OF STATE INVOLVING 'HAVING'. THE ENTAILMENT FRAME '(NOW) A HAS B' IS SUGGESTED FOR 'C GIVES A B'. 'NOW' IS NOT POSITED AS A DESCRIPTIVE CONSTANT, HOWEVER. INSTEAD, 'AT TIME T' IS DEFINED [P. 63] AND, FOR 'GIVE', 'A HAS B AFTER TIME T' IS USED. [P. 76] THE RESULT IS THAT NO FORMAL, EXPLICIT DESCRIPTION OF TIME OR TENSE HAS BEEN GIVEN. BENDIX ADMITS AS MUCH ON P. 77, VIZ. "IT IS IMPORTANT TO POINT OUT THAT TENSES HAVE NOT BEEN SPECIFICALLY CONSIDERED IN THIS STUDY. A FULLER DESCRIPTION OF A LANGUAGE WOULD, OF COURSE, HAVE TO GIVE A SEMANTIC ANALYSIS OF THE SYSTEM OF TENSES AS WELL AS STATE THE COMBINATORIAL SEMANTIC RULES THAT ACCOUNT FOR THE TEMPORAL, ASPECTUAL, AND OTHER INTERACTIONS BETWEEN THE VERBAL MEANINGS AND THE TENSE SYSTEM." A MORE SERIOUS DEFECT IS BENDIX' FAILURE TO PERCEIVE THAT THE CAUSAL CONNECTION BETWEEN THE ACTION OF THE GIVER AND THE TRANSFER OF POSSESSION TO THE RECIPIENT MUST INCLUDE A SPECIFICATION OF THE GIVER'S INTENTION. SEE CONCEPTUAL ANALYSES I-101 ('CAUSE') AND I-103 ('INTENTIONAL ACTION'). AN ANALYSIS CONTAINING 'HAS' AS A COMPONENT MUST BE PREPARED TO FACE THE COMPLEXITIES ARISING WHEN ABSTRACT OR NON-PHYSICAL OBJECTS ARE "GIVEN". THUS, 'HAS' MUST BE GIVEN A NON-POSSESSIVE INTERPRETATION IN SENTENCES SUCH AS THE FOLLOWING. "SHE GAVE ME A SHY SMILE" (CP. W7 DEFINITION OF SENSE 5B OF 'GIVE': "TO PRESENT TO VIEW OR OBSERVATION").

Semanal #: S-165

Source: DIXON, R.M.W. [1973] "THE SEMANTICS OF GIVING"

Accompanying Words: DONATE, PRESENT, AWARD, PAY, LEND, SELL, RENT

Author's Sense: "[THIS FIRST SENSE OF 'GIVE' INDICATES] GIVING IN RETURN FOR SOME SERVICE OR OBJECT, NORMALLY ACCORDING TO A DEFINITE CONTRACT OR AGREEMENT. THUS: [1] FRED GAVE TOM \$2 FOR CUTTING THE LAWN. [2] JONES GAVE HIS MEN 50% EXTRA FOR WORKING ON SUNDAY. [3] SALLY GAVE JOHN THE CLOCK FOR \$5. [4] JOHN GAVE SALLY \$5 FOR THE CLOCK." [P. 211]

W7 Sense: VT 10A: TO YIELD POSSESSION OF BY WAY OF EXCHANGE: PAY

Predicate-Args: /GIVE1/ [SUBJECT] [OBJECT] [TO INDIRECT OBJECT] [FOR PHRASE EXPRESSING THE CONTRACT]

Informal Explic.: "NUCLEAR VERB 'GIVE' HAS SYNTACTIC CONNECTION WITH S[UBJECT], O[BJECT], TO I[NDIRECT] O[BJECT] OBLIGATORILY; FOR EXC[HANGE] OPTIONALLY." [P. 222]

Informal Explic.: "IN THE PRESENCE OF...A FOR-PHRASE, THE INDIRECT OBJECT MAY BE OMITTED:...JOHN GAVE \$5 FOR THE BOOK. ...'GIVE' CANNOT OCCUR WITHOUT A DIRECT OBJECT; WE DO NOT GET: JOHN GAVE FOR THE BOOK." [P. 213]

Semanal #: S-166

Source: DIXON, R.M.W. [1973] "THE SEMANTICS OF GIVING"

Accompanying Words: DONATE, PRESENT, AWARD, PAY, LEND, SELL, RENT

Author's Sense: "[THIS SECOND SENSE OF 'GIVE' INDICATES] SPONTANEOUS

30 April 1974

System Development Corporation
TM-5243/001/00

23

GIVING, WHICH IS NOT IN DIRECT CONTRACTUAL SETTLEMENT OF ANY OBLIGATION (ALTHOUGH I MAY OF COURSE BE INDIRECTLY MOTIVATED BY SOME GOOD DEED THAT THE RECIPIENT HAS AT SOME TIME PERFORMED). THUS: [5] TOM GAVE \$100 TO THE HOSPITAL. [6] MARY GAVE EFFIE HER OLD FUR COAT." [P. 211]

W7 Sense: VT 1: TO MAKE A PRESENT OF

2A: TO GRANT OR BESTOW BY FORMAL ACTION

3A: TO PUT INTO THE POSSESSION OF ANOTHER FOR HIS USE

3C: TO COMMIT TO THE TRUST OR KEEPING OF ANOTHER

Predicate-Arqs: /GIVE2/ [SUBJECT] [OBJECT] [INDIRECT OBJECT]

Informal Explic.: "IN THE SENSE OF SPONTANEOUS GIVING, 'GIVE2' REQUIRES BOTH DIRECT AND INDIRECT OBJECTS; NEITHER CAN BE DELETED. ADMITTEDLY THERE ARE SPECIAL CIRCUMSTANCES IN WHICH ONE OR BOTH OF THESE DOES NOT OCCUR PER SE IN SURFACE STRUCTURE --FOR INSTANCE 'JOHN ALWAYS GIVES GENEROUSLY ON FLAG DAYS'; BUT NOTE THAT IN THIS SENTENCE 'GENEROUSLY' GIVES SOME INFORMATION ABOUT THE DEEP DIRECT OBJECT AND 'ON FLAG DAYS' ENABLES ONE TO INFER THE NATURE OF THE INDIRECT OBJECT. OUR POINT IS THAT GIVEN ANY SIMPLE SENTENCE INVOLVING 'GIVE2', NEITHER DIRECT NOR INDIRECT OBJECT CAN AUTOMATICALLY BE DELETED." [P. 213]

Semanal #: S-167

Source: KATZ, J. [1972] SEMANTIC THEORY

Accompanying Words: BUY, SELL, TRADE, EXCHANGE, SWAP, RECEIVE, LET, HIRE, RENT, LEND, BORROW, CHARTER

Author's Sense: "...A PROCESS IN WHICH THE POSSESSION OF SOMETHING IS TRANSFERRED FROM ONE PERSON TO ANOTHER WITHOUT ANYTHING BEING GIVEN IN RETURN. ...[7.176] [A] JOHN GAVE THE BOOK TO MARY" [P. 348]

W7 Sense: VT 1: TO MAKE A PRESENT OF

2A: TO GRANT OR BESTOW BY FORMAL ACTION

3A: TO PUT INTO THE POSSESSION OF ANOTHER FOR HIS USE

3C: TO COMMIT TO THE TRUST OR KEEPING OF ANOTHER

Prelim. Qual.: BY REMOVING FROM THE DEFINITION OF "SELL" THE PROCESS SEMANTIC MARKER WHICH INDICATES THE SUM OF MONEY FOR WHICH AN ITEM IS SOLD, THE LEXICAL READING FOR "GIVE" CAN BE OBTAINED. [PARAPHRASE OF KATZ ON P. 248]

Predicate-Arqs: 'GIVE': (SUBJECT-OP) (OBJECT-OP) (INDIRECT-OBJECT-OP) (INFLXIONAL)

Components: ((CONDITION) (POSSESSES Y) OF X AT T-I)

((CONDITION) (POSSESSES Y) OF Z AT T-J)

Component Composition: THE PROCESS OF GIVING IS REPRESENTED BY A TRANSITION FROM THE STATE REPRESENTED BY THE FIRST COMPONENT TO THE STATE REPRESENTED BY THE SECOND COMPONENT. 'X' REPRESENTS THE SUBJECT, 'Y' THE OBJECT, AND 'Z' THE INDIRECT OBJECT. 'T-I' INDICATES THE INITIAL STATE AND 'T-J' THE TERMINAL STATE.

Selectional Features: 'Y' = <(PHYSICAL OBJECT)>

'X' = <(HUMAN) & (NOT-INFANT)>

'Z' = <(HUMAN) & (NOT-INFANT)>

SOLAP Comments: KATZ' DEFINITION HAS THE OUTSTANDING QUALITY OF REPRESENTING THE PROCESS INVOLVED IN GIVING. IT IS DEFICIENT IN OTHER RESPECTS, HOWEVER. FIRST, IT FAILS TO INDICATE THAT THE SUBJECT ('X') INITIATES/CAUSES THE PROCESS. SECOND, IT DOES NOT INDICATE THAT IN THE TERMINAL STATE (T-J) THE SUBJECT (X) NO LONGER POSSESSES THE OBJECT (Y).

3.2.3 Explanatory Notes on Descriptive Constants

To aid the user in determining the significance of the analyses proposed, the archive includes a set of notes explaining as precisely as possible the author's intention in using a particular descriptive constant or non-standard symbol. The notes following are those needed for the semantic analysis of 'give' by Bendix, which is found in the preceding section.

Constant: CAUSE(S)

Explan #: E-113

Source: BENDIX, EDWARD [1966]

Definition: "WE USE THE LABEL 'CAUSE' FOR THE COMPONENT IN QUESTION HERE WHEN TESTS INDICATE THE PRESENCE OF THE COMPONENT THAT 'A' (OR 'I', ETC.) DOES OR IS SOMETHING AND, FURTHER, THAT THIS ACTION OR STATE OF 'A' LEADS TO, RESULTS IN, OR MAY OTHERWISE BE SAID TO BE CONNECTED WITH ANOTHER ACTION OR STATE IN A WAY THAT WE MAY CALL CAUSAL. ULTIMATELY WE WOULD WANT TO DEFINE THE METALINGUISTIC TERM 'CAUSE' MORE PRECISELY. THIS WOULD BE INTENDED AS A DEFINITION NOT OF CAUSATION, BUT OF WHAT WE CLAIM A SPEAKER IS ASSERTING WHEN HE USES A FORM WHOSE MEANING WE DEFINE AS CONTAINING THE COMPONENT 'CAUSE'." [P. 63]

SOLAR comments: IT APPEARS TO BE THE CASE THAT THE CONSTANT 'CAUSE' IS BEING USED TO EXPRESS THE NOTION 'BRING ABOUT BY INTENTIONAL ACTION'. SEE THE SOLAR COMMENTS IN SEMANTIC ANALYSIS S-102 AND THE CONCEPTUAL ANALYSES OF 'CAUSE (I-101) AND 'INTENTIONAL ACTION' (I-103).

Word Analyzed: GIVE, TAKE

Semanal #: S-102, S-103

Constant: A- (AN-)

Explan #: E-114

Source: BENDIX, EDWARD [1966]

Definition: "THE EXISTENTIAL QUANTIFIER 'THERE IS A...' IS INDICATED IN A DEFINITION BY PREFIXING THE INDEFINITE ARTICLE "A- (AN-)" TO THE FIRST OCCURRENCE OF THE QUANTIFIED VARIABLE IN THE DEFINITION." [P. 14] "THUS, THE ABBREVIATED NOTATION 'B AN-R' A' MAY ALSO BE READ 'THERE IS A RELATION SUCH THAT B IS IN THE RELATION TO A AND THE RELATION IS A MEMBER OF THE SUBCLASS OF H-RELATIONS'. FURTHERMORE, "A (AN)" WILL BE PREFIXED ONLY TO THE FIRST OCCURRENCE OF A QUANTIFIED VARIABLE IN A DEFINITION, IN IMITATION OF THE OBJECT LANGUAGE. IN THE INFORMAL REPHRASINGS OF DEFINITIONS, "SOME" WILL OCCASIONALLY BE USED FOR THIS "A (AN)" WHEN A GIVEN REPHRASING APPEARS THEREBY TO BE REINDEXED MORE NATURALLY OR FAITHFULLY.

Word Analyzed: GIVE, TAKE

Semanal #: S-102, S-103

30 April 1974

25

System Development Corporation
TM-5243/001/00

Constant: A, B, C,...

Explan #: E-115

Source: BENDIX, EDWARD [1966]

Definition: THE LETTERS "A, B, C..." REPRESENT VARIABLES OR THE SYNTACTIC POSITIONS OF NOUN PHRASE. IT IS TO SIMPLIFY OUR DESCRIPTION THAT WE USE THESE LETTERS AS SYNTACTIC SYMBOLS RATHER THAN THE MORE USUAL NOTATION NP1, NP2, NP3,...OR NP, NP', NP'',..... WHEN THESE LETTERS ARE ENCLOSED IN SINGLE QUOTES, EITHER ALONE (E.G., 'A') OR AS PART OF A LONGER STRIP (E.G., 'A HAS B'), THEY REPRESENT REFERENTS OF TOKENS OF TYPES OCCURRING IN THE GIVEN NOUN-PHRASE POSITIONS...." [13]

Word Analyzed: GIVE, TAKE

Semanal #: S-102, S-103

Constant: HAS (PH)

Explan #: F-116

Source: BENDIX, EDWARD [1966]

Definition: "THE BASIC COMPONENT OF THE MEANING OF 'A HAS B' HAS BEEN DEFINED AS 'THERE IS A RELATION BETWEEN A AND B'. THE SUBCLASS OF RELATIONS TO WHICH THIS RELATION MAY BELONG WAS DEFINED FURTHER IN CHAPTER 3 IN TERMS OF A NUMBER OF PARAPHRASING CONSTRUCTIONS OR SOURCE-SENTENCE TYPES. BRIEFLY THEN, USING 'H' TO INDICATE THE RECIPROCAL RELATION OF 'HAVE': 'A HAS B' = 'B AN-RH A' (TO BE READ: 'B IS IN AN H-RELATION TO A' OF 'THERE IS AN H-RELATION BETWEEN B AND A') WHERE: 'B AN-RH A' REPRESENTS THE CLASS OF PARAPHRASE OR SOURCE CONSTRUCTIONS OF 'A HAS B'" [P. 62]

Word Analyzed: GIVE, TAKE

Semanal #: S-102, S-103

Constant: AN-F

Explan #: E-118

Source: BENDIX, EDWARD [1966]

Definition: "'AF' SYMBOLIZES A FUNCTION WITH VARIABLE 'A'. WE PLACE THE SYMBOL FOR THE VARIABLE BEFORE THE SYMBOL FOR THE FUNCTION, RATHER THAN IN THE MORE CUSTOMARY ORDER 'F(X)' OR 'FX', TO INDICATE CORRESPONDENCE WITH THE SUBJECT-PREDICATE CONSTRUCTION (OR WITH THE TOPIC-COMMENT FORM)." [P. 13] IN THE INFORMAL EXPLANATION OF 'GIVE', 'C AN-F' IS TAKEN AS 'C DOES C IS SOMETHING'. [P. 62]

Word Analyzed: GIVE

Semanal #: S-102

Constant: R

Explan #: E-119

Source: BENDIX, EDWARD [1966]

Definition: "'ARB', WITH VARIABLES 'A' AND 'B', STANDS FOR A RELATION BETWEEN 'A' AND 'B'." [P. 13] "[THE DEFINITION OF] 'B AN-R C' [IS] ('B IS IN SOME RELATION TO C' OR 'THERE IS A RELATION BETWEEN B AND C'-- I.E., THE SUBCLASS OF RELATIONS IS NOT SPECIFIED)" [P. 62]

Word Analyzed: GIVE, TAKE

Semanal #: S-102, S-103

Constant: '

Explan #: E-120

Source: BENDIX, EDWARD [1966]

Definition: "SINGLE QUOTES ARE USED HERE TO ENCLOSE MEANINGS OR ELEMENTS OF MEANING SUCH AS COMPONENTS, DEFINITIONS, GLOSSES, TRANSLATIONS, OR INTERPRETATIONS. UNDERLINING IS FREQUENTLY EMPLOYED FOR INDICATING FORMS, SYNTACTIC ELEMENTS, OR THE DEFINIENDUM OF A DEFINITION." [P. 13]

SOLAR comments: Since underlining is not possible in SOLAR, we have replace 'Bendix' underline with single quotes around the underscored word(s). We believe no significant ambiguities have been introduced by this convention.

Word Analyzed: GIVE, TAKE

Semanal #: S-102, S-103

3.2.4 Conceptual Analyses of the Notions 'Cause', 'Intentional Action', and 'Responsible'

The analyses of 'cause' and 'intentional action' are directly relevant to Bendix's semantic analysis of 'give' (see S-102 in Section 3.2.2). The analysis of 'responsible' is included here because it has close conceptual connections with the other two, as indicated under its "qualifications" heading. Such cross-references will subsequently be reinforced by more detailed cross-references within the analyses proper when a sufficiently large number of conceptual analyses have been prepared so that such interconnections can be set up in a systematic way.

30 April 1974

27

System Development Corporation
TM-5243/001/00

Notion: CAUSE

Constants: CAUSE, CAUSE1

Words: GIVE [S-102], TAKE [S-103], DRIVE [S-129, S-130], DROP [S-132]

Inteq. Sum #: I-101

Source: LEWIS, DAVID [1973] "CAUSATION", THE JOURNAL OF PHILOSOPHY,
PP.

556-567.

KIM, JAEGWON [1973] "CAUSES AND COUNTERFACTUALS", THE JOURNAL
OF PHILOSOPHY, PP. 570-572.

LEWIS, DAVID [1973B] COUNTERFACTUALS (BLACKWELL AND HARVARD)
ANSCOMBE, G. L. M. [1971] CAUSALITY AND DETERMINATION.
CAMBRIDGE

MACKIE, J. L. [1965] "CAUSES AND CONDITIONS" AMERICAN
PHILOSOPHICAL QUARTERLY, PP. 245-264.

WHITE, MORTON, FOUNDATIONS OF HISTORICAL KNOWLEDGE [HARPER AND
ROW, 1965] PP. 14-104.

W7 Sense: 1A: SOMETHING THAT OCCASIONS OR EFFECTS A RESULT

Qualifications: THE NOTION OF 'CAUSE' ANALYZED HERE IS THAT OF ONE
PARTICULAR EVENT BEING A CAUSE OF ANOTHER PARTICULAR EVENT.

Analysis: "IN THIS CASE...THE SO CALLED CAUSE IS, AND IS KNOWN TO BE, AN
INSUFFICIENT BUT NECESSARY PART OF A CONDITION WHICH IS ITSELF
UNNECESSARY BUT SUFFICIENT FOR THE RESULT." [MACKIE, J. L.
1965]. LEWIS STATES THE REGULARITY ANALYSES OF CAUSATION AS
FOLLOWS: "LET C* BE THE PROPOSITION THAT C EXISTS (OR OCCURS)
AND LET E* BE THE PROPOSITION THAT E EXISTS. THEN C CAUSES E,
ACCORDING TO A TYPICAL REGULARITY ANALYSIS, IF [1] C* AND E*
ARE TRUE; AND [2] FOR SOME NON-EMPTY SET L\$ OF TRUE
LAW-PROPOSITIONS AND SOME SET P\$ OF TRUE PROPOSITIONS OF
PARTICULAR FACT, L\$ AND P\$ JOINTLY DO NOT IMPLY E*, AND P\$
ALONE DOES NOT IMPLY C*-->E*.... IT REMAINS TO BE SEEN WHETHER
ANY REGULARITY ANALYSIS CAN SUCCEED IN DISTINGUISHING GENUINE
CAUSES FROM EFFECTS, EPIPHENOMENA, AND PREFEMPTED POTENTIAL
CAUSES--AND WHETHER IT CAN SUCCEED WITHOUT FALLING VICTIM TO
WORSE PROBLEMS, WITHOUT PILING ON THE EPICYCLES, AND WITHOUT
DEPARTING FROM THE FUNDAMENTAL IDEA THAT CAUSATION IS
INSTANTIATION OF REGULARITIES." (LEWIS [1973] PP. 556-557; IN
TRANSCRIBING THIS PASSAGE WE USED A '\$' ADSCRIPT TO INDICATE
SCRIPT CAPITAL LETTERS IN THE ORIGINAL, AND A '*' ADSCRIPT TO
INDICATE BLOCK CAPITALS .) LEWIS THEN GOES ON TO GIVE A
MODIFIED SINE QUA NON ANALYSIS: "CAUSAL DEPENDENCE AMONG ACTUAL
EVENTS IMPLIES CAUSATION. IF C AND E ARE TWO ACTUAL EVENTS
SUCH THAT E WOULD NOT HAVE OCCURRED WITHOUT C, THEN C IS A
CAUSE OF E. BUT I REJECT THE CONVERSE. CAUSATION MUST ALWAYS
BE TRANSITIVE; CAUSAL DEPENDENCE MAY NOT BE; SO THERE CAN BE
CAUSATION WITHOUT CAUSAL DEPENDENCE. LET C, D, AND E BE THREE
ACTUAL EVENTS SUCH THAT D WOULD NOT HAVE OCCURRED WITHOUT C AND
E WOULD NOT HAVE OCCURRED WITHOUT D. THEN C IS A CAUSE OF E

EVEN IF E WOULD STILL HAVE OCCURRED (OTHERWISE CAUSED) WITHOUT C. WE EXTEND CAUSAL DEPENDENCE TO A TRANSITIVE RELATION IN THE USUAL WAY. LET C, D, E, ... BE A FINITE SEQUENCE OF ACTUAL PARTICULAR EVENTS SUCH THAT D DEPENDS CAUSALLY ON C, E ON D, AND SO ON THROUGHOUT. THEN THIS SEQUENCE IS A CAUSAL CHAIN. FINALLY, ONE EVENT IS A CAUSE OF ANOTHER IFF THERE EXISTS A CAUSAL CHAIN LEADING FROM THE FIRST TO THE SECOND. THIS COMPLETES MY COUNTERFACTUAL ANALYSIS OF CAUSATION." LEWIS [1973] p. 563.

Critique: FOR CRITICISM OF THE LEWIS ANALYSIS, SEE KIM [1973]. THE WORST CASE FOR LEWIS IS THAT OF AN EVENT B THAT IS A SUBEVENT OF AN EVENT A. FOR A SOPHISTICATED REGULARITY ANALYSIS, SEE MACKIE [1965]. FOR ANOTHER DISCUSSION OF THE REGULARITY THEORY PLUS A DISCUSSION OF HOW TO ANALYSE 'THE CAUSE OF' SEE WHITE [1965]. FOR DISCUSSION OF THE NEED TO ALLOW CAUSATION UNDER INDETERMINISM SEE ANSCOMBE [1971].

Notion: INTENTIONAL ACTION

Constants: CAUSE, CAUSE1

Words: GIVE [S-102], TAKE [S-103], DRIVE [S-129, S-130], DROP [S-132]

Semant. #: S-102, S-103, S-129, S-130, S-132

Inteq. Sum. #: I-103

Source: ARMSIRONG, D. M. [1971] "ACTING AND TRYING", UNPUBLISHED PRESIDENTIAL ADDRESS AUSTRALASIAN ASSOCIATION OF PHILOSOPHY.

CHISHOLM, R. M. "THE DESCRIPTIVE ELEMENT IN THE CONCEPT OF ACTION", JOURNAL OF PHILOSOPHY, 1964, PP. 614-625.

CHISHOLM, R. M. [1970] "THE STRUCTURE OF INTENTION", JOURNAL OF PHILOSOPHY, PP. 633-647.

DAVIDSON, D. [1963] "ACTIONS, REASONS, AND CAUSES", JOURNAL OF PHILOSOPHY, PP. 685-700.

OLDENQUIST, ANDREW [1967] "CHOOSING, DECIDING, AND DOING" IN THE ENCYCLOPEDIA OF PHILOSOPHY, 2, PP. 96-104.

W7 Sense: 2: THE BRINGING ABOUT OF AN ALTERATION BY FORCE OR THROUGH A NATURAL AGENCY

Qualifications: THE NOTION OF 'INTENTIONAL ACTION' ANALYZED HERE PRESUPPOSES THE NOTIONS OF 'TRYING', 'ACTION' AND 'CAUSE'. IT IS CLOSELY LINKED TO SUCH OTHER NOTIONS AS 'INTENTIONS', 'GOAL' AND 'RESPONSIBLE'.

Analysis: IN "ACTING AND TRYING", D. M. ARMSTRONG GIVES THE FOLLOWING ANALYSIS OF "A INTENTIONALLY PERFORMS ACTION P": A DID P INTENTIONALLY IF AND ONLY IF [1] A DID P [2] A TRIED TO DO P [3] A'S TRYING TO DO P CAUSED A TO DO P. AS ARMSTRONG NOTES (18), THIS DEFINITION IS STILL TOO WEAK: MOVEMENTS NOT INTUITIVELY CONSIDERED INTENTIONAL ACTIONS SATISFY [1]-[3]. AN EXAMPLE BASED ON AN INSIGHT OF CHISHOLM, "THE DESCRIPTIVE ELEMENT IN THE CONCEPT OF ACTION", SHOWS WHY. ARMSTRONG PUTS IT AS FOLLOWS: "A SETS OUT ACROSS TOWN WITH THE OBJECT OF KILLING HIS UNCLE. THE EXCITEMENT GENERATED BY THIS PLAN CAUSES HIM TO DRIVE FAST AND CARELESSLY. AS A RESULT HE KNOCKS DOWN AND KILLS A PEDESTRIAN WHOM HE DOES NOT SEE. IT TURNS OUT TO BE A UNCLE. A'S ATTEMPT TO MURDER HIS UNCLE BROUGHT IT ABOUT THAT HE KILLED HIS UNCLE. BUT HIS KILLING WAS NOT INTENTIONAL." ARMSTRONG GOES ON TO SUGGEST THAT THE ANALYSIS NEEDS TO BE FILLED OUT SO THAT A'S TRYING TO DO P BRINGS ABOUT P ACCORDING TO A CAUSAL PATTERN INVOLVING A'S PURPOSES AND BELIEFS THAT CAN BE SHADOWED OUT IN A TRAIN OF PRACTICAL REASONINGS. WE CAN FORMALIZE THIS AS FOLLOWS: A DOES P INTENTIONALLY IF AND ONLY IF [1] A DOES P BY DOING Q [2] A TRIES TO DO P BY DOING Q [3] A'S TRYING TO DO P BY DOING Q CAUSES A TO DO Q [4] IF A'S DOING Q CAUSES EVENT E, AND A'S DOING P CONSISTS IN CAUSING E, THEN A'S DOING Q CAUSES E ACCORDING TO A CAUSAL CHAIN THAT WAS PARTIALLY ANTICIPATED BY A, AND A'S PARTIALLY ANTICIPATING C CAUSED A TO DO (OR TRY TO DO) P BY DOING Q. (THIS IS TO BE UNDERSTOOD AS INCLUDING THE POSSIBILITY THAT Q IS P). IN THE ABOVE EXAMPLE, IT SHOULD BE CLEAR THAT THOUGH A KILLED HIS UNCLE BY HITTING HIM WITH A CAR, CLAUSES [2]-[4] FAIL SINCE A DIDN'T TRY TO KILL HIS UNCLE BY HITTING HIM WITH A CAR. SUPPOSE, HOWEVER, THAT [1] A INTENTIONALLY KILLS HIS UNCLE BY SHOOTING A GUN. THEN PRESUMABLY, [2] HE WILL HAVE TRIED TO KILL THE UNCLE BY SHOOTING THE GUN. AND [3] THAT ATTEMPT TO KILL HIS UNCLE BY SHOOTING THE GUN WILL BE THE REASON FOR (AND HENCE, CAUSE OF) HIS SHOOTING THE GUN. BUT A'S SHOOTING THE GUN CAUSES HIS UNCLE TO DIE AND A'S KILLING HIS UNCLE CONSISTS IN A'S CAUSING HIS UNCLE TO DIE, SO IF [4] IS TO BE SATISFIED THEN A'S SHOOTING THE GUN MUST CAUSE HIS UNCLE TO DIE ACCORDING TO A CAUSAL CHAIN C THAT WAS PARTIALLY ANTICIPATED BY A (NAMELY, THE BULLET HITTING THE UNCLE AND CAUSING DEATH), AND A'S ANTICIPATING THAT CAUSAL CHAIN MUST BE A REASON FOR (AND HENCE A CAUSE OF) A'S KILLING THE UNCLE BY SHOOTING THE GUN. THAT NOT JUST ANY CAUSAL CHAIN FROM SHOOTING THE GUN TO THE UNCLE'S DYING IS SUFFICIENT FOR A TO HAVE INTENTIONALLY KILLED HIS UNCLE CAN BE SEEN BY CONSIDERING THE POSSIBILITY THAT THE BULLET MISSES BUT THE EPISODE IS SO DNNERVING TO A'S UNCLE THAT HE HAS A HEART ATTACK AND DIES. IF A HAS INTENDED THAT HE KILL HIS UNCLE BY THE BULLET PIERCING THE UNCLE'S BODY, THEN, IN THIS CASE IT SEEMS THAT A DID NOT INTENTIONALLY KILL HIS UNCLE. IF, ON THE OTHER HAND, A HAD THOUGHT, "WELL, EVEN IF THE BULLET MISSES MAYBE THE OLD GEEZER WILL HAVE A HEART ATTACK AND DIE", THEN A WOULD HAVE INTENTIONALLY KILLED HIS UNCLE.

Critique: THERE ARE STILL SOME DIFFICULTIES WITH THIS ANALYSIS. THE PRACTICAL REASONING OR CAUSAL CHAIN ANTICIPATED BY A COULD PERHAPS BE WRONG IN SOME DETAILS (WHILE SUBSTANTIALLY CORRECT) IN A CASE OF INTENTIONAL ACTION, THOUGH THIS ANALYSIS DOESN'T DEAL WITH SUCH CASES. IT SHOULD BE NOTED THAT THE ANALYSIS AUTOMATICALLY YIELDS A NOTION OF 'AGENT OF INTENTIONAL ACTION': A "IS THE AGENT OF" THE INTENTIONAL ACTION P IF AND ONLY IF A IS THE ONE WHO DOES P INTENTIONALLY. FOR A DEFENSE OF REASONS AND INTENTIONS AS CAUSES OF ACTIONS, SEE DAVIDSON [1963]. FOR MORE ON 'INTENTION', SEE CHISHOLM [1970]. A SURVEY ARTICLE, DEALING WITH RELATIONS BETWEEN ACTION, CHOICE AND DECISION IS OLDENQUIST, [1969].

Notion: RESPONSIBLE

Constants: RESPONSIBLE

Words: EXCUSE [S-101]

Inteq. Sum #: I-102

Source: CHISHOLM, R. [1967] "'HE COULD HAVE DONE OTHERWISE'", THE JOURNAL OF PHILOSOPHY, PP. 409-417.

FRANKFURT, HARRY [1969] "ALTERNATE POSSIBILITIES AND MORAL RESPONSIBILITY", THE JOURNAL OF PHILOSOPHY, DECEMBER 4, PP. 829-839.

FRANKFURT, HARRY [1971] "FREEDOM OF THE WILL AND THE CONCEPT OF A PERSON", THE JOURNAL OF PHILOSOPHY, JANUARY 14, PP. 5-20.

HOLBORN, L. C. [1972] "BLAME, PRAISE, AND CREDIT", PROCEEDINGS OF THE ARISTOTELIAN SOCIETY, 1971-72, PP. 85-100.

HOOK, SIDNEY [1966] "NECESSITY, INDETERMINISM, AND SENTIMENTALISM", IN FREE WILL AND DETERMINISM (ED. BY B. BEROFKY). HARPER & ROW, N.Y., PP. 26-45.

KAUFMAN, ARNOLD [1967] "RESPONSIBILITY, MORAL AND LEGAL", IN THE ENCYCLOPEDIA OF PHILOSOPHY (ED. BY P. EDWARDS). THE MACMILLAN CO. & THE FREE PRESS, NEW YORK, PP. 183-188.

PASKE, GERALD [1970] "RESPONSIBILITY AND THE INCOMPARABILITY PRINCIPLES", THE PERSONALIST, PP. 477-485.

w7 Sense: 1A: "LIABLE TO BE CALLED UPON TO ANSWER AS THE PRIMARY CAUSE, MOTIVE, OR AGENT".

Qualifications: THE NOTION OF 'RESPONSIBLE' ANALYZED HERE HAS CLOSE CONCEPTUAL CONNECTIONS WITH NUMEROUS OTHER PHILOSOPHICALLY CENTRAL NOTIONS, E.G., 'INTENTION', 'INTENTIONAL ACTION', 'ACTION', 'CONSCIOUSNESS', 'RATIONALITY', 'FREEDOM', 'CAUSE', 'FREEWILL', 'FREE AGENT', 'AGENT', 'DETERMINISM', AND 'JUSTICE'; THE ENTRIES FOR THESE NOTIONS SHOULD BE CONSULTED.

Analysis: "IF WE MAKE A LIST OF THE CIRCUMSTANCES BEHIND ACTIONS FOR WHICH WE HOLD INDIVIDUALS RESPONSIBLE AND THOSE FOR WHICH WE DO NOT, WE SHALL FIND THAT AS A RULE THE FIRST CLASS CONSISTS OF THOSE IN WHICH PRAISE AND REWARD, BLAME AND PUNISHMENT, TEND TO INFLUENCE THE FUTURE CONDUCT OF THOSE INVOLVED AND/OR THOSE TEMPTED. THIS IS NOT THE WHOLE STORY.... THE BEHAVIOR OF INFANTS...IS MODIFIABLE BY APPROPRIATE REWARD AND PUNISHMENT EVEN THOUGH WE DO NOT HOLD THEM MORALLY RESPONSIBLE. BUT AS THE AGE OF RATIONALITY APPROACHES, WE GRADUALLY DO. THIS SUGGESTS THAT IN ADDITION TO SUSCEPTIBILITY TO REWARD AND PUNISHMENT, WE ATTRIBUTE RESPONSIBILITY WHERE THERE IS A TENDENCY TO RESPOND TO VALID REASONS, TO BEHAVE RATIONALLY, TO RESPOND TO HUMAN EMOTIONS IN A HUMAN WAY. PERHAPS A THIRD ELEMENT INVOLVED IN THE ATTRIBUTION OF MORAL RESPONSIBILITY TO VOLUNTARY ACTION IS THE ASSUMPTION THAT VOLUNTARY ACTION IS APPROVED ACTION. A MAN IS MORALLY RESPONSIBLE FOR AN ACTION HE COMMITS TO THE EXTENT THAT HE APPROVES OF IT. IF HE SINCERELY DISAPPROVES OF HIS ACTION, REGARDS IT AS WRONG AND CONDEMNS IT AS WRONG, BUT STILL COMMITS IT, WE TEND TO REGARD HIM AS ILL, AS ACTING UNDER "COMPULSION". IT IS SOME SUCH CONSIDERATIONS AS THIS THAT LIES BEHIND OUR EXTENUATION OF CERTAIN KINDS OF APPARENTLY VOLUNTARY ACTION (AS WHEN WE SAY, 'HE DIDN'T MEAN TO DO IT'), ESPECIALLY WHERE IGNORANCE IS PRESENT" HOOK, SIDNEY [1966], PP. 47-48. THE RELATIONSHIP BETWEEN COERCION AND MORAL RESPONSIBILITY HAS RECENTLY BEEN CLARIFIED BY FRANKFURT. MOST PHILOSOPHERS HAVE AGREED THAT WHAT FRANKFURT CALLS "THE PRINCIPLE OF ALTERNATE POSSIBILITIES" IS TRUE, VIZ., THAT A PERSON IS MORALLY RESPONSIBLE FOR WHAT HE HAS DONE ONLY IF HE COULD HAVE DONE OTHERWISE. HOWEVER, FRANKFURT ADDUCES EXAMPLES THAT SHOW CONCLUSIVELY THAT THE PRINCIPLE OF ALTERNATE POSSIBILITIES IS FALSE. HE OBSERVES, "THE FACT THAT A PERSON COULD NOT HAVE AVOIDED DOING SOMETHING IS A SUFFICIENT CONDITION OF HIS HAVING DONE IT. BUT, AS SOME OF MY EXAMPLES SHOW, THIS FACT MAY PLAY NO ROLE WHATEVER IN THE EXPLANATION OF WHY HE DID IT. IF SOMEONE HAD NO ALTERNATIVE TO PERFORMING A CERTAIN ACTION BUT DID NOT PERFORM IT BECAUSE HE WAS UNABLE TO DO OTHERWISE, HE WOULD HAVE PERFORMED EXACTLY THE SAME ACTION EVEN IF HE COULD HAVE DONE OTHERWISE. THE FOLLOWING MAY ALL BE TRUE: THERE WERE CIRCUMSTANCES THAT MADE IT IMPOSSIBLE FOR A PERSON TO AVOID DOING SOMETHING; THESE CIRCUMSTANCES ACTUALLY PLAYED A ROLE IN BRINGING IT ABOUT THAT HE DID IT, SO THAT IT IS CORRECT TO SAY THAT HE DID IT BECAUSE HE COULD NOT HAVE DONE OTHERWISE; THE PERSON REALLY WANTED TO DO WHAT HE DID; HE DID

IT BECAUSE IT WAS WHAT HE REALLY WANTED TO DO, SO THAT IT IS NOT CORRECT TO SAY THAT HE DID WHAT HE DID ONLY BECAUSE HE COULD NOT HAVE DONE OTHERWISE. UNDER THESE CONDITIONS, THE PERSON MAY BE MORALLY RESPONSIBLE FOR WHAT HE HAS DONE. ON THE OTHER HAND, HE WILL NOT BE MORALLY RESPONSIBLE FOR WHAT HE HAS DONE IF HE DID IT ONLY BECAUSE HE COULD NOT HAVE DONE OTHERWISE, EVEN IF WHAT HE DID WAS SOMETHING HE REALLY WANTED TO DO." FRANKFURT, HARRY [1969]

Critique: HOOK'S CLAIM THAT "A MAN IS MORALLY RESPONSIBLE FOR AN ACTION HE COMMITS TO THE EXTENT THAT HE APPROVES OF IT" IS SUSPECT, AT LEAST PENDING SOME FURTHER EXPLANATION OF 'APPROVAL'. IF HOOK IS USING 'APPROVAL' IN A NON-TECHNICAL SENSE, THEN IT WILL FOLLOW THAT ONE CAN EVADE RESPONSIBILITY FOR ONE'S ACTIONS MERELY BY DISAPPROVING OF THEM. THE TRUTH THAT IS BEHIND HOOK'S POSITION IS THAT DISAPPROVAL IS EVIDENCE THAT THE ACTION IS NOT REALLY ONE'S OWN. THE "COMPULSION" OF WHICH HOOK SPEAKS CAN POSSIBLY BE EXPLAINED VIA A SPLITTING UP OF THE INDIVIDUAL INTO PARTS, SOME ONE OF WHICH--THE ONE WITH WHICH THE SELF IS IDENTIFIED--BEING THE PART WHICH DISAPPROVES OF THE ACTION. A POSSIBLE APPROACH IS SKETCHED IN FRANKFURT [1971] WHERE A DISTINCTION BETWEEN FIRST-ORDER VOLITIONS AND SECOND-ORDER VOLITIONS (VOLITIONS ABOUT FIRST-ORDER VOLITIONS) IS PUT TO GOOD USE. FRANKFURT MENTIONS BUT DOES NOT GO INTO THE COMPLEXITIES OF, HIGHER THAN SECOND-ORDER VOLITIONS. A DETAILED ANALYSIS OF 'HE COULD HAVE DONE OTHERWISE' IS GIVEN BY CHISHOLM [1967] AND AN INTERESTING TREATMENT OF THE RELATION BETWEEN RESPONSIBILITY AND RATIONALITY IS GIVEN BY PASKE [1970]. RELEVANT TO FILLMORE'S USE OF 'RESPONSIBLE (SEE S-101) BUT TOO DETAILED TO BE INCLUDED HERE, IS HOLBOROW [1972]. A SUMMARY OF THE RELATIONSHIP BETWEEN MORAL AND LEGAL RESPONSIBILITY IS GIVEN BY KAUFMAN [1967].

3.2.5 Other Activities

The bibliographic file and the word index have been implemented in a data management system, and listings have been produced periodically.

In the area of program development, we have completed writing, compiling, and debugging the programs needed to restructure the W7 parsed transcripts into a format suitable for input to other programs that will produce the various machine-derived files. One of these latter programs (the one building the collocational feature file) has been written, compiled, and debugged.

We have publicized the archive throughout the US, Canada, Australia, and Europe. Approximately 20 researchers have responded to our solicitation of documents dealing with lexical semantics, and about 35 have expressed interest in receiving data from the archive.

3.3 PLANS

During the remainder of this contract year, LDA will be focusing on five tasks:

- 1) Continue data collection from the literature. This will involve extending the bibliographic file by about 700 entries, updating the word index, and more than doubling the semantic and conceptual analysis files.
- 2) Continue to develop the large number of programs needed to produce the three remaining machine-derived files.
- 3) Put through production runs to build the machine-derived files.
- 4) Disseminate data from each of the files.
- 5) Document the archive as it comes into existence.

During the next contract year, emphasis will be placed on updating and improving each of the files, adding antonymitive relations to the archive, and facilitating distribution of data from the archive. The updating activity derives from the continual addition of words to the SUR lexicons. Improvements include restructuring based on feedback from users regarding the utility of each file. Antonymitive relationships will be added to the semantic field file to permit their incorporation into the semantic networks of the SUR systems. To facilitate use of the archive, the project will produce user's guides, provide demonstrations of the use of SOLAR, and build a user data-management interface program to guide the user in performing on-line retrieval.

3.4 STAFF

Dr. T. C. Diller, Project Leader
J. Olney (Consultant)
F. Heath (part-time)

3.5 DOCUMENTATION AND PUBLICATIONS

- [1] T. Diller, and J. Olney, "SOLAR: A Semantically-Oriented Lexical Archive," SDC SP-3726, NIC 19927, SUR Note 110, November 1973.
- [2] T. Diller, "SOLAR Bibliography User's Guide," SDC TM-5292, in press.

4. COMMON INFORMATION STRUCTURES

4.1 INTRODUCTION

The Common Information Structures project is addressing the problem of converting and transferring data bases among disparate data management systems (DMSs). The need to share data for different applications, as well as the need to transfer existing data to be used by new computer systems, makes it apparent that general techniques for data base conversion are desirable. The goal of this project is to develop techniques for data base conversion that are practical for application to current data management systems and that are designed to be easily used by data base users.

The difficulties in converting a data base from one data management system (DMS) to another arise from the fact that data base structures are system and application dependent. As a result, the DMS imposes constraints on the form of the data base. These constraints are of three types: (1) logical-level constraints, such as level of hierarchies, size and number of fields, and data types; (2) storage-level constraints, such as inversion and access path of files and file indexing organization; and (3) physical-level constraints, such as physical devices used and block/record structures. These levels were described in reports issued for the 1972-73 contract year.

The conventional method of converting data bases for new applications is to write a special-purpose conversion program for each data base. Another possible approach is to define data description languages for all three levels mentioned above, then specify in these languages the source and target data bases, as well as conversion statements between them.¹ Since this approach involves all three levels, it requires complex and detailed data description languages, which are difficult to learn and to use. It also requires that data be converted from the source physical environment to the corresponding target physical environment, which further complicates any possible implementation.

¹This approach has been discussed by several researchers. See D. P. Smith, "An Approach to Data Description and Conversion" (Univ. of Penn. Ph.D. dissertation, 1971); J. A. Ramirez, "Automatic Generation of Data Conversion Programs Using a Data Description Language" (Univ. of Penn. Ph.D. dissertation, 1973); and three papers published in Proceedings of the 1972 ACM SIGFIDET Workshop (New York: Association for Computing Machinery): "An Approach to Stored Data Definition and Translation" (Stored Data Definition and Translation Task Group, pp. 13-56), "A Method for Data Translation Using the Stored Data Definition and Translation Task Group Languages" (D. P. Smith, pp. 107-124), and "A Developmental Model for Data Translation" (J. P. Fry, R. L. Frank, and E. A. Hershey, III, pp. 77-106).

The approach being taken by this project is based on the idea that the data conversion process can depend mainly on conversion at the logical level. Conversion at this level can be achieved by using existing query and generate capabilities of DMSs to move data from their physical representation to the logical level and vice versa. The tasks required in the data conversion process are diagrammed in Figure 4-1. First, the source data base is retrieved using the query capabilities of the source DMS and reformatted into a standard form. Then, the data translation process takes place, and a target data base in the standard form is produced and reformatted into a data format acceptable to the generate capability of the target DMS. Finally, the target data base is generated with the generate capability of the target DMS.

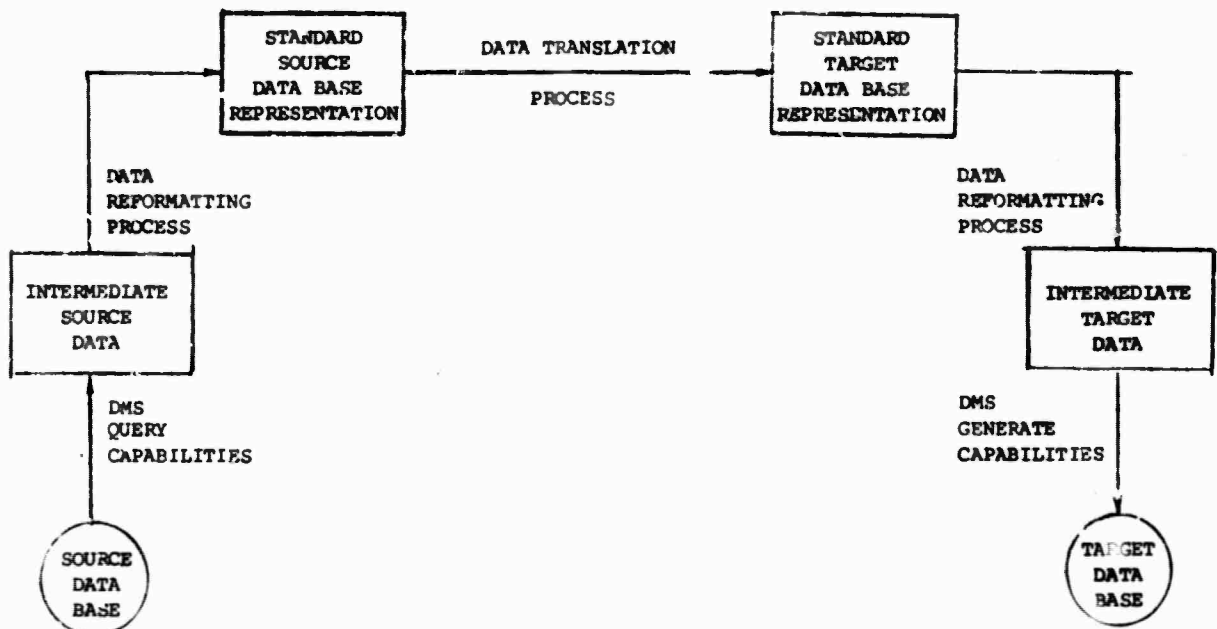


Figure 4-1. Data Flow in the Data Conversion Process

4.2 PROGRESS AND PRESENT STATUS

The first task of this project during the present contract year was to explore the possibilities of automatically generating target data descriptions from source data descriptions. The file definition languages (FDLs) of several DMSSs were studied in an attempt to isolate the restrictions these systems impose on data base structures. After a thorough analysis of these FDLs, it was concluded that automatic generation of a target description from the source description is either:

- 1) Trivial--when the restrictions on the complexity of data base structures of the target system are less constraining than the restrictions of the source system (for example, going from a system that allows one level of hierarchy into a system that allows nine levels of hierarchy); or
- 2) Impossible to predict in the general case, because the target description is semantically dependent on the intended use of the data base and on time, space, and cost considerations.

Consequently, we concluded that it would not be fruitful to continue in this direction, and we decided to direct the main effort of the project at the development of processes that actually convert data, rather than at the automatic generation of target data descriptions. In the context of developing the details of the data conversion process, the following tasks were performed:

- 1) A detailed study was made of the different data description languages and data conversion approaches described in the literature.¹ We concluded that these approaches advocate the use of details of all three levels of data description and are therefore equally difficult to implement and use. This led us to the current approach, which depends mainly on the logical level of data.
- 2) A methodology for the data conversion process was developed. It involves the development of source and target reformatters and a logical data translator. These processes are driven by statements written in three languages, which are dependent mainly on logical characteristics of the data to be converted. These languages are the Common Data Description Language (CDDL), the Common Data Translation Language (CDTL), and the Common Data Format Language (CDFL).

¹ See E. H. Sibley and R. W. Taylor, "A Data Definition and Mapping Language," *Comm. ACM*, December 1973 (pp. 750-759); R. W. Taylor, "Generalized Data Base Management System Data Structures and their Mapping to Physical Storage" (Univ. of Michigan Ph.D. dissertation, 1971); D. P. Smith, "An Approach to Data Description and Conversion" (loc. cit.); and CODASYL Systems Committee, "Feature Analysis of Generalized Data Base Management Systems" (New York: Association for Computing Machinery, 1971).

- 3) A preliminary version of CDDL was developed. This was accomplished after an analysis of existing DDLs (especially those of CODASYL and Smith) and selecting the subset that is relevant to our approach. The syntax of the current version of CDDL is given in Section 4.2.1. The syntax of CDDL represents a logical view of hierarchical data structures. Because it contains no storage-level or physical-level requirements, it is a simple language to be specified by a user. At the same time, the language is flexible enough to represent a given data management system's limitations (such as hierarchy levels) for any data base organized in this DMS. The statements in CDDL, together with statements in CDTL, supply all the necessary information for the data conversion process. As a consequence, CDDL might change according to the requirements of CDTL.
- 4) The functions necessary for the Common Data Translation Language (CDTL) were developed. These functions are represented in terms of conversion statements between fields of the source data description and the target data description. A conversion statement consists of an association of one source field (or more) to a target field, plus an algorithm for the conversion of data (such as truncation, concatenation, or data-type transformation). Types of conversion statements were identified and form the basis of CDTL. A summary of these types is given in Section 4.2.2.

4.2.1 Common Data Description Language (CDDL) Specification

CDDL is a simple language. A file statement consists of group and field statements. A group statement, in turn, consists of field statements and, possibly, additional group statements, thus establishing a hierarchical structure. There are several types for fields, and a repetition number for groups puts an upper limit on the number of values in a group instance.

```
datadescription = filestatement {fieldstatement | groupstatement}
filestatement = FILE <filename : recordlength : filegroupname>
fieldstatement = FIELD <fieldname : datatype>
groupstatement = GROUP <groupname : grouplength : {<FIELD : fieldname> |
                                <GROUP : groupname : repetition>}>
filename = {character}
recordlength = integer | NOLIMIT
filegroupname = groupname
fieldname = {character}
datatype = string | number | picture
```

```

groupname = {character}
grouplength = integer | NOLIMIT
repetition = integer | NOLIMIT
string = C (stringlength)
number = I (int) | FP (fixedparameters) | FL (floatparameters)
picture = PICT pictspec
integer = {digit}
stringlength = integer
int = integer
fixedparameters = int . int
floatparameters = fixedparameters E int
pictspec = '{# | @ | * | character}'
digit = 1|2|3|4|5|6|7|8|9|0
character = A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z|
           a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|z|
           0|1|2|3|4|5|6|7|8|9|,|.|?|!|
           (|)|_|$|-|+|_|/|=|&|~|
           " ' | * | @ | " < | " > | " " | : | " #

```

4.2.2 Types of Data Conversion Functions

After analyzing file description languages of existing DMSSs, we concluded that the following conversion functions would be most useful:

- 1) Instance mapping--represents a mapping of instances of a field of a repeating group (RG) into a field of a higher level RG.
- 2) Inverse instance mapping--represents a combination of multiple values of fields of the same RG level into an RG instance of a lower level.
- 3) Operator mapping--allows a set of values in an RG instance to be combined by some operation (e.g., Average) into one value of a field in a higher-level RG.
- 4) Simple field mapping--allows for an association of source and target fields according to a given algorithm (e.g., truncation).
- 5) Field value repetition--necessary when a repetition of a field value through values of a lower-level RG is required.

- 6) Level creation--can be used to create a target RG level according to a criterion on a source RG.
- 7) Field partition--allows the splitting of an RG into two or more RGs (e.g., an employee RG into an exempt employee RG and a non-exempt employee RG).
- 8) Field conjunction--the inverse function to field partition.
- 9) Field concatenation--necessary when a target field is made up by concatenating source fields (or portions of them).
- 10) Intermediate level elimination--can be used for the elimination of an RG level present in the source but unnecessary in the target data base.
- 11) Inversion--necessary when an alternate view of the data base is required (e.g., a department-employee data base needs to be re-organized as an employee-department data base).

4.3 PLANS

During the remainder of the contract year, the design and specification of CDTL and the design of the logical data translator will be completed. Six tasks are involved:

CDTL Design and Specification

- 1) An analysis of the semantic implications of data base conversions to determine the desirable combination of functions in CDTL. This will eliminate the possibility of a data conversion requirement that is semantically impossible.
- 2) The development of a specification of CDTL according to the functions required.
- 3) Design of the standard data format to be used by the data translator and the reformatters.

Logical Data Translator Design

- 4) Design of a process that will perform a lexical and semantic analysis of the source and target data base descriptions in CDDL and of the translation statements in CDTL and produce internal tables that represent the data translation operations to be performed.

30 April 1974

40

System Development Corporation
TM-5243/001/00

- 5) Design of read/write modules that can input data and generate output data.
- 6) Design of control modules that restructure source data instances into target data instances by utilizing the read/write modules and the tables produced by the lexical and semantic analysis process.

By the end of this contract year, the project will be ready to begin implementation of the conversion functions according to the CDDL and CDTL statements. In parallel, we plan to study the functions necessary for CDFL and develop it.

4.4 STAFF

A. Shoshani, Project Leader
K. J. Fogt